Bayesian inference for high dimensional factor copula models

by

Hoang Nguyen

in partial fulfillment of the requirements for the degree of Doctor in Business and Quantitative Methods

Universidad Carlos III de Madrid

Advisor(s):

María Concepción Ausín Olivera

Pedro Galeano San Miguel

March 25, 2019

To my family and friends!

Acknowledgements

My Ph.D. study is a cheerful journey with friends and family. I would like to express my love to those who have accompanied with me along the road.

First of all, I would like to thank Prof. M.Concepción Ausín & Prof. Pedro Galeano for their excellence guidance, patience, and their endless support during my PhD. They have taught me many things, especially the way of thinking, the way of conveying the new ideas, and the way of organizing the workflow. They have made my learning curve become smoother. They have encouraged me throughout different stages of my training process. Without them, the thesis cannot be done.

I would like to thank Prof. J.Miguel Marín who first introduced me to Bayesian Statistics. Since my master dissertation, his suggestion of Stan software had opened the door to the Bayesian world. My idea can be quickly prototyped and tested in Stan before I improve it. Juanmi is also my dear neighbor who I can share and listen to his advice about my personal life.

I am especially grateful to Prof. Michael Wiper for his insightful comments and suggestions. Mike is a wisdom tree who always can clear my doubts. His immense knowledge of Bayesian Statistics has helped me improve my work and saved me from the maze of prior distributions.

I am also very grateful to Prof.Roberto Casarin for his hospitality during my exchange at Ca' Foscari University of Venice. I was extremely happy to work with Prof. Roberto. From my original proposal, he had suggested several developments and modifications. His advice kept questioning me on how to improve and extend my current work.

I would like to express my gratitude to Audrone Virbickaite & Huong Nguyen who helped me

since the first day I came to Spain. Their instructions and recommendations led me to Bayesian inference and pursuit the doctoral program in Statistics. I am contented with my Bayesian choice.

I would like to thank friends and colleagues from Universidad Carlos III de Madrid. Thank Javier, my office mate, for his generous conversations and sharing about life and hobbies. Thank Nicolas for all the moments and cares. Thank Angela, Maria, Mario, Antonio Vázquez, Cristina, Faiza for pulling me out of work and making me less homesick. Thank Prof. Eduardo, Iñaki, David, and Antonio from the Coding Club UC3M for remarkable and professional work. Thank Prof. Helena Veiga for the lectures and advice. Thank Prof. Stefano Cabras for the unforgettable experience at ISBA conference. Thank Susana, Paco, Almudena for their administration services.

Last but not least, I would like to thank my parents and my brother for their unconditional love. You always support me no matter of rainy or sunny days. I miss you very much. Thank Yen, my girlfriend, who have cooked the best dishes. After a tired day, I know that I can still keep smiling with your everlasting stories. I love you with all my heart.

Thank you all for standing with me. “Your speed doesn’t matter, forward is forward”.

PUBLISHED AND SUBMITTED CONTENT

- H. Nguyen, M. C. Ausín, and P. Galeano. Parallel Bayesian inference for high dimensional dynamic factor copulas. *Journal of Financial Econometrics*, 17(1):118–151, 2019.
  - Coauthor;
  - https://doi.org/10.1093/jjfinec/nby032
  - The paper is included in **Chapter 2** of the thesis;
  - The material from this source included in this thesis is not singled out with typographic means and references.
- H. Nguyen, M. C. Ausín, and P. Galeano. Variational inference for high dimensional structured factor copulas. *UC3M Working Papers Statistics and Econometrics*, WP18-05, 2018.
  - Coauthor;
  - https://e-archivo.uc3m.es/handle/10016/27652
  - The paper is included in **Chapter 3** and one section in **Chapter 4** of the thesis;
  - The material from this source included in this thesis is not singled out with typographic means and references.

Abstract

Copulas have been applied to many research areas as multivariate probability distributions for non-linear dependence structures. However, extending copula functions in high dimensions is challenging due to the increase of model parameters and computational intensity. Fortunately, in many circumstances, high dimensional dependence can be explained by a few common factors. This dissertation focuses on using factor copula models to analyze the high dimensional dependence structure of random variables. Different factor copula models are proposed as a solution for the curse of dimensionality. Then, a parallel Bayesian inference or a Variational Inference (VI) is employed to speed up the computation time. Chapter 2 concentrates on a dynamic one factor model for group generalized hyperbolic skew Student-t copulas. Chapter 3 and 4 extend the multi-factor copula models to suit with different high dimensional data sets. These models have applications in a wide variety of disciplines, such as financial stock returns, spatial time series, and economic time series, among others.

Chapter 2 develops a class of dynamic one factor copula models for tackling the curse of dimensionality. The asymmetric dependence is taken into account by group generalized hyperbolic skew Student-t copulas. The study is influenced by Creal and Tsay [2015], Oh and Patton [2017b], but instead, the dynamic factor loading equation follows a generalized autoregresive score process which depends on the copula density conditional on the factor rather than the unconditional copula density, as proposed in Oh and Patton [2017b]. As the conditional posterior distributions of parameters in groups can be inferred independently due to model specifications, a parallel Bayesian inference is employed. This reduces the time of computation for a sizable problem from several days to one hour using a personal computer. The model is illustrated for 140 firms listed in

the S&P500 index and the optimal portfolio allocation is obtanied based on minimum Conditional Value at Risk (CVaR). The major content of this chapter resulted into a paper by Nguyen et al. [2019] which had been accepted for publication in *Journal of Financial Econometrics*.

Chapter 3 takes advantage of the static structured factor copula models proposed by Krupskii and Joe [2015a] for the dependence of homogenerous variables in different groups. To extend one factor copula models, Krupskii and Joe [2015a] assume a hierarchical structure for the latent factors and model the dependence of the observables through a serial of bivariate copula links between the observables and the latent variables. This topology stems from vine copulas and becomes very flexible to capture both asymmetric tail dependence as well as correlation among variables. The VI is used to estimate the different specifications of structured factor copula models. VI aims to approximate the joint posterior distribution of model parameters by a simpler distribution which resuls in a very fast inference algorithm in comparison to the MCMC approach. Secondly, an automated procedure is proposed to recover the dependence linkages. By taking advantage of the posterior modes of the latent variables, the initial assumptions of bivariate copula functions are inspected and replaced for better copula functions based on the Bayesian information criterion (BIC). Chapter 3 shows an empirical example where the structured factor copula models help to predict the missing temperatures of 24 locations among 479 stations in Germany. The major content of this chapter resulted into a working paper by Nguyen et al. [2018].

Chapter 4 supplements the factor copula model with a combination of a factor copula model at the first tree level and a truncated vine copula structure at a higher tree level. The model is not only suitable to capture different behaviors at the tail of the distribution but also remains parsimonious with interpretable economic meanings. The truncated factor vine copula models can outperform the multi-factor copula model in cases that there is weak dependence among variables in higher tree levels and the inference of group latent factors becomes inaccurate. The VI strategy is used and the dependence structure can be recovered with a similar copula selection procedure. Chapter 4 compares the statistical criteria of different factor models for the dependence structure of stock returns from 218 companies listed in 10 different European countries.

Contents

List of Figures

Chapter 1

Introduction

Copulas have become an essential tool for modelling non-standard multivariate distributions as they allow for skewness and fat tails in the marginal distributions and a non-linear dependence structure, see Cherubini et al. [2011], Patton [2012] and Fan and Patton [2014], among others. Although the idea of copula was developed by Sklar [1959], it became popular among scholars at the end of the nineties due to the development of quantitative risk management methodology, see Embrechts [2009]. Copulas are preferred over the classical multivariate distributions as among other aspects, they allow more parameters to control for the tail dependence. With a few time series, standard copula families such as the elliptical and the Archimedean copulas are usually applied. However, when the dimension increases, the use of these standard copula functions is problematic. For instance, the Student-t copula is only able to fit well in small dimentions, see Demarta and McNeil [2005] and Creal and Tsay [2015]. Also in many empirical datasets, asymmetric dependence is often found in the lower tail and upper tail of the joint distribution.

To extend the copula models in high dimensions, Bedford and Cooke [2001, 2002], Aas et al. [2009] propose vine copulas and Krupskii and Joe [2013], Oh and Patton [2017a] come up with factor copulas. In vine copula models, the dependence structure of variables is constructed as a graphical object linked by a serial of bivariate copula functions and conditional bivariate copula functions between observables. In factor copula models, a few latent variables are assumed to

affect each random variable, and conditional on these latent variables, observable variables become independent. Each approach has its own strength and weakness. Vine copula models can capture well the correlation as well as the tail dependence. However, the number of parameters becomes explosive when the number of dimensions increases which results in a truncated vine copula at some levels, see Brechmann et al. [2012]. Moreover, Morales-Nápoles [2010] addresses that there are huge possibilities of regular tree vines that could be used. Alternatively, factor copula models are proposed to prevent the curse of dimensionality as the number of parameters will scale linearly with the number of dimensions. Adding or subtracting variables does not change the dependence structure. However, the latent factors make it difficult to estimate and perform the model selection. We introduce the construction and properties of copulas before taking a deeper analysis of factor copula models in the following sections.

1.1 Copula definition

The copula notation was first introduced by Sklar [1959] as an alternative approach for modelling the joint distribution of random variables. Copulas allow us to separate the marginal distributions from the dependence structure and incoportate more parameters to control for the tail dependence in comparison to the classical multivariate distributions. Smith [2011] considers copulas as an easier way of modelling dependence by switching from the domain of the data to the unit hypercube.

Sklar [1959]'s Theorem: Let $X = (X_1, \ldots, X_d)'$ be the d-dimensional random variable where the joint cumulative distribution function (cdf) is F, and the marginal distributions are $F_1, \ldots, F_d$. There exists a copula function C, such that

$$F(x_1, \ldots, x_d) = C(F_1(x_1), \ldots, F_d(x_d)) = C(u_1, \ldots, u_d) \tag{1.1}$$

where $u_i = F_i(x_i)$ for $i = 1, \ldots, d$. Sklar's Theorem is one of the most important results of copulas since it shows that a distribution function can be written in terms of a copula function in the transformed unit domain. If the variables have continuous marginal distributions, the copula

function is unique. Note that the probability integral transformation, $U_i = F_i(X_i)$, has uniformly distributed marginals in $[0, 1]$. The joint probability density function of a multivariate distribution, f, can be written as a product of a copula density function c and marginal densities, $f_1, ..., f_d$, by taking derivative of Eq 1.1 :

$$f(x_1, ...x_d) = c(u_1, \ldots, u_d) \times \prod_{i=1}^{d} f_i(x_i). \tag{1.2}$$

This separation relaxes the restriction between different marginal distributions and the joint distribution function, hence copula models can be applied effectively in many research areas. Furthermore, the estimation of copula parameters could be done in two stages, see Joe [2005] and Chen and Fan [2006]. The first stage aims to estimate the parameters θ_i of the marginal distributions and obtain an approximate sample of the copula observations $u_i = F_i(x_i|\bar{\theta}_i)$ for $i = 1, \ldots, d$ where $\bar{\theta}_i$ is an estimator of θ_i using the maximum likelihood or the Bayesian approach. Then, the second stage will account for the copula parameters based on the pseudo- observables $u = \{u_1, \ldots, u_d\}$. There are a large number of bivariate copula functions which is suitable with different types of data, see Joe [1997]. However, it is difficult to extend the bivariate copula functions to trivariate and higher dimensions. The next section describes the most important bivariate copula families which are used to construct complex copula functions in high dimensions.

1.2 Bivariate copula families

Elliptical copulas and Archimedean copulas are the most well-known copula families that are easily derived and they are capable of wide ranges of dependence. However, these bivariate copula families only have one or two parameters controlling the correlation as well as the tail dependence, therefore they are more appropriate with small and medium sample sizes.

1.2.1 Elliptical copulas

Elliptical copulas are constructed based on the elliptical distributions. The Gaussian copula and the Student-t copula are commonly used members of the elliptical copula family. They inherits good properties of the elliptical distribution such as a similar form of the conditional distribution and joint distribution function. It is also straightforward to extend the bivariate copula to higher dimension using the multivariate elliptical distributions. The simulation of elliptical copulas can be easily carried out from inverting elliptical distributions to unit domains. However, elliptical copulas do not have a closed form expression. For example, a bivariate Gaussian copula function is

$$C_{Gauss}(u_1, u_2) = \int_{\infty}^{\Phi^{-1}(u_1)} \int_{\infty}^{\Phi^{-1}(u_2)} \frac{1}{2\pi(1-\rho)^{1/2}} \exp\left\{-\frac{s_1^2 - 2\rho s_1 s_2 + s_2^2}{2(1-\rho)}\right\} ds_1 ds_2$$

where $\rho \in [-1, 1]$ and Φ is the standard normal distribution function.

Figure 1.1 shows the contour plots of bivariate elliptical copulas with the same standard normal marginal. Gaussian copulas shows no tail dependence while Student-t copulas has symmetric tail dependence. For that reason, Student-t copulas are more suitable for heavy tail dependence.

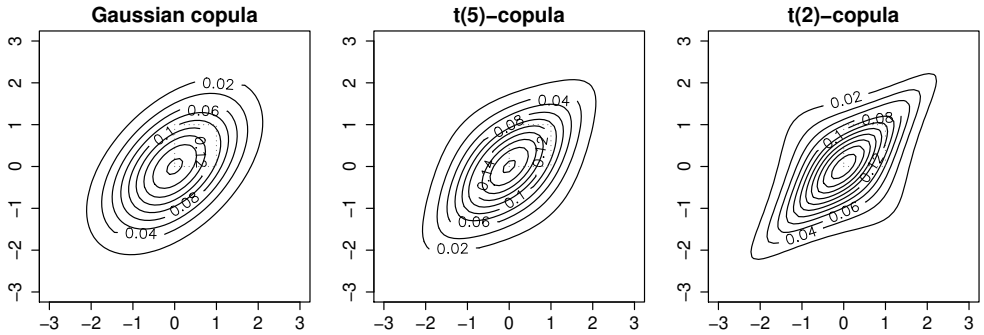

Figure 1.1: Contours of bivariate elliptical copulas with the same standard normal marginal

1.2.2 Archimedean copulas

Different from elliptical copulas, Archimedean copulas allow for asymmetric tail behaviors to cope with different types of data. Figure 1.2 shows different contour plots of Archimedean copulas with

the same standard normal marginal. Clayton copula has lower tail dependence while Gumbel copula has upper tail dependence and Frank copula shows no tail dependence. It is also easy to rotate Archimedean copulas for $90°$, $180°$, $270°$ to create rotated bivariate Archimedean copula functions.

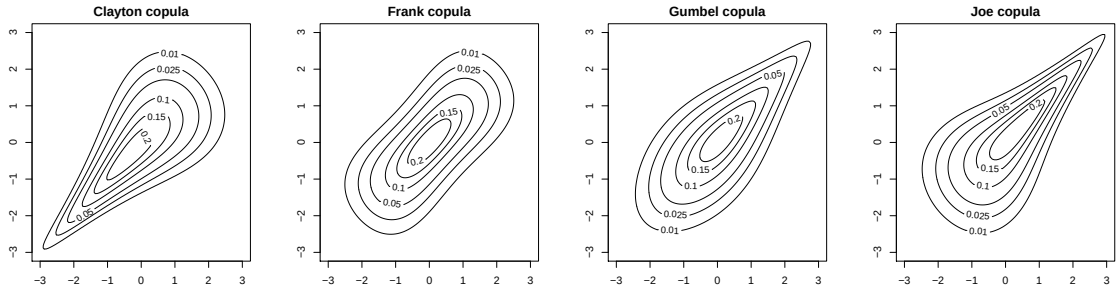

Figure 1.2: Contours of bivariate Archimedean copulas with the same standard normal marginal

Another advantage of Archimedean copulas is that the copula functions can be written in a closed form expressions. For example,

$$C_{Clayton}(u_1, u_2) = \left(u_1^{-\theta} + u_2^{-\theta} - 1\right)^{-\frac{1}{\theta}} \text{ where } \theta \in (0, \infty),$$
$$C_{Gumbel}(u_1, u_2) = \exp\left[-\left\{(-\log u_1)^{\theta} + (-\log u_2)^{\theta}\right\}^{1/\theta}\right] \text{ where } \theta \in [1, \infty).$$

The bivariate Archimedean copulas are constructed based on a generator function φ such that φ is a continuous, strictly decreasing convex function from $[0, 1]$ to $[0, \infty]$ where $\varphi(1) = 0$. The *pseudo-inverse* of φ is the function $\varphi^{[-1]} : [0, \infty] \mapsto [0, 1]$. The Archimedean copula function is defined by its generator function and pseudo-inverse function,

$$C(u_1, u_2) = \varphi^{[-1]}(\varphi(u_1) + \varphi(u_2)).$$

Archimedean copulas could be extended for multivariate Archimedean copulas, see McNeil et al. [2010] or for hierarchical Archimedean copulas, see Savu and Trede [2010] and Okhrin et al. [2013]. Multivariate Archimedean copulas restrict the rank correlation matrix to be equal as there

is only one parameter that controls for the dependence. Hierarchical Archimedean copulas allow for more flexibility, but they suffer the intensive computation.

1.3 Vine copula

There are several approaches to extend the bivariate copulas to high dimensions. Joe [1994, 1996] propose a D-vine copula for multivariate extreme value distributions. A D-vine copula for d variables are defined recursively through $d(d-1)/2$ bivariate copulas and its conditional distributions. Independently, Bedford and Cooke [2001, 2002] develop a general definition of vine copulas. Aas et al. [2009] derive an algorithm for making inferences of vine parameters. The vine copulas are built from a serial of bivariate copulas and conditional bivariate copulas which extend the flexibility of dependence structure among the variables. For example, the density for a 3 dimensional vine copulas could be written as

$$\begin{aligned} f(x_1, x_2, x_3) &= f(x_3|x_1, x_2) f(x_2|x_1) f(x_1) \\ &= \Big[c_{13|2}(F(x_3|x_2), F(x_1|x_2)) f(x3|x_2)\Big] \Big[c_{12}(F(x_1), F(x_2)) f(x_2)\Big] f(x_1) \\ &= \Big[c_{13|2}(F(x_3|x_2), F(x_1|x_2))\Big] \Big[c_{32}(F(x_3), F(x_2))\Big] \Big[c_{12}(F(x_1), F(x_2))\Big] \prod_{i=1}^{d=3} f(x_i). \end{aligned}$$

Of course that there are different ways to decompose variables, see Aas et al. [2009]. Among those, C-vine and D-vine structures are two commonly used vine copulas. In C-vine copula, each tree has a unique node at the root that connects to all other nodes, while in the D-vine copulas, each tree is a path. More applications of vine copulas could be found in Joe and Kurowicka [2010].

In general, the combinations of bivariate copula functions and dependence tree in vine copulas make them very flexible to capture different dependence patterns in the middle as well as in the tail of the joint distribution. Vine copula models are preferred in low and medium dimensions. However, there are several disadvantages. Firstly, there are arguments on the choice of tree structure. The bivariate copula linkages are sensitive to the selected structure so adding variables

can change the current structure completely. Dissmann et al. [2013] propose a heuristic algorithm to identify layer structure of the tree vine sequentially. Secondly, the number of parameters increase as a square function of dimension which makes estimation very expensive. However, truncated vine copula at some levels help to reduce the estimated paramters of vine copula models, see Brechmann et al. [2012].

1.4 Factor copulas

Factor copula models assume that the observed variables are conditionally independent given one or more latent variables. Factor copulas models have been used previously in the literature as a solution for the curse of dimensionality. For instance, Hull and White [2004] propose a model based on combining linearly the common factor risk and idiosyncratic risk for valuing tranches of collateralized debt obligations and n^{th} to default swaps. Andersen and Sidenius [2004], and van der Voort [2007] improve the model by considering a non-linear factor structure while Murray et al. [2013] extend for multi-factor Gaussian copulas.

There are mainly two approaches to set up factor copula models. Krupskii and Joe [2013, 2015a] propose pair copula construction-based factor models while Creal and Tsay [2015] and Oh and Patton [2017b] extend the classical factor analysis by inverting the dependence structure from latent elliptical or skew-elliptical distributions to the constrained copula domain. Krupskii and Joe [2013, 2015a] construct a general class of factor copulas where the dependence structure is decomposed into a sequence of bivariate copulas and conditional bivariate copulas between variables and the latent factors. Hence, factor copulas could be considered as a truncated C-vine copula with latent variables. Specifically, if bivariate Gaussian linking copulas are used, then the factor copula model can be seen as a copula version of the multivariate Gaussian distribution where the correlation matrix has a factor structure. Otherwise, if bivariate non-Gaussian linking copulas are used, the model is able to describe tail asymmetry and tail dependence. However, it is difficult to extend the models to the dynamic settings. Alternatively, Creal and Tsay [2015] and Oh

and Patton [2017b] incorporate the class of dynamic factor models proposed in the literature of time series analysis with arbitrary marginal distributions. Therefore, the dependence structure of variables could be the same as the classical dynamic factor models but with arbitrary marginal distributions. Nevertheless, the choice of copula functions is limited to some extensions of elliptical distributions such as the Student-t and the skew Student-t distributions. In the thesis, Chapter 2 follows the Creal and Tsay [2015] and Oh and Patton [2017b] approach while Chapter 3 and Chapter 4 extends the Krupskii and Joe [2013, 2015a] approach.

Besides the mentioned applications in finance, factor copulas also have been applied to different types of datasets, for example, spatial dependence of temperatures in Krupskii et al. [2016], spatio-temporal dependence of hourly wind data in Krupskii and Genton [2017], mortality dependence of multiple populations in Chen et al. [2015], behavior dependence of item response in Nikoloulopoulos and Joe [2015], and extreme dependence of river flows in Lee and Joe [2017].

1.5 Dependence measures

Rank correlations and tail dependence are two common benchmarks for non-linear dependence. Rank correlations are preferred over the linear correlation because it is invariant to variable transformations. As one of the properties of copulas is that the dependence structure remains unchanged when using a monotonic transformation of variables, rank correlations and tail dependence can be written in terms of copula functions.

1.5.1 Rank correlations

There are several measures of rank correlation such as Kendall's τ, Spearman's ρ, and Blomqvist's β, see Joe [2014]. Among those, Kendall's τ is the most frequently used as Kendall's τ takes into

account the difference between the concordance and discordance of bivarite $(U_1, U_2)'$, as follows:

$$\begin{aligned}\rho_\tau(U_1, U_2) &= P((U_1 - \hat{U}_1)(U_2 - \hat{U}_2) > 0) - P((U_1 - \hat{U}_1)(U_2 - \hat{U}_2) < 0) \\ &= 4 \iint_{[0,1]^2} C(u_1, u_2) dC(u_1, u_2) - 1,\end{aligned}$$

where $(\hat{U}_1, \hat{U}_2)'$ is an independent copy of $(U_1, U_2)'$. The rank correlations of common bivariate copula functions are shown in Table 3.1. Also, Spearman's ρ is commonly used as a measures of the linear correlation between U_1 and U_2,

$$\rho_S(U_1, U_2) = 12 \iint_{[0,1]^2} (C(u_1, u_2) - u_1 u_2) du_1 du_2$$

If the random variables U_1 and U_2 are completely independent, $\rho_\tau(U_1, U_2) = \rho_S(U_1, U_2) = 0$.

1.5.2 Tail dependence

Tail dependence of copula models is one of the main concerns in empirical applications. For example, in risk management, if a stock return reduces more than 5%, what is the probability that other stock returns also reduces correspondingly? Tail dependence measures the dependence in the upper right quadrant or lower left quadrant. The coefficient of lower tail and upper tail dependence of two variables U_1 and U_2 are defined respectively as

$$\begin{aligned}\lambda_L &= \lim_{u \to 0} P(U_2 \leq u | U_1 \leq u) = \lim_{u \to 0} \frac{C(u, u)}{u}, \\ \lambda_U &= \lim_{u \to 1} P(U_2 > u | U_1 > u) = 2 + \lim_{u \to 0} \frac{C(1-u, 1-u) - 1}{u}.\end{aligned}$$

Here, U_1 and U_2 are asymptotically independent in the lower tail (upper tail) if $\lambda_L = 0$ ($\lambda_U = 0$). The measurements of tail dependence for a bivariate copula model can be calculated asymptotically or at a quantile, see McNeil et al. [2010], however it is difficult to compare with those implied by empirical data due to limited sample sizes. Krupskii and Joe [2015b] propose the tail-weighted

dependence as the correlation of transformed variables in which puts heavier weights for the extreme observations,

$$\rho_L = Cor\left(\alpha\left(1-\frac{U_1}{p}\right), \alpha\left(1-\frac{U_2}{p}\right)\middle| U_1 < p, U_2 < p\right),$$
$$\rho_U = Cor\left(\alpha\left(1-\frac{1-U_1}{p}\right), \alpha\left(1-\frac{1-U_2}{p}\right)\middle| 1-U_1 < p, 1-U_2 < p\right)$$

where $p \leq 0.5$ and $\alpha(\cdot)$ is a continuous increasing function. The application of tail dependence will be analyzed deeper for different factor copula models in the next chapters.

1.6 Overview of the thesis

Each chapter of the thesis extends the factor copula models for high dimensional datasets and seeks for the solution of the computational issues. Then, several applications of factor copula models are illustrated in finance and economic contexts.

Chapter 2 focuses on a class of dynamic one factor copula models where the dynamic factor loading equation depends on the copula density conditional on the factor rather than the unconditional copula density, as proposed in Oh and Patton [2017b]. The model also accounts for the asymmetric dependence with group generalized hyperbolic skew Student-t copulas. The conditional posterior distributions of parameters in groups can be inferred independently due to model specifications. Hence, a parallel Bayesian inference is employed to reduce the computation burden. Chapter 2 shows an example of portfolio allocation and risk management of 140 firms listed in the S&P500 index. The major content of the this chapter resulted into a paper by Nguyen et al. [2019].

Chapter 3 analyses the static structured factor copula models proposed by Krupskii and Joe [2013, 2015a]. Alternative to the frequentist approach in the original paper, Chapter 3 applies VI to estimate the different specifications of structured factor copula models. VI aims to approximate the joint posterior distribution of model parameters by a simpler distribution, hence it speeds up the computational time in comparison to the MCMC approach. Another issue of factor copula models is that the bivariate copula functions connecting the variables are unknown in high dimensions.

An automated procedure is derived to recover the dependence structure. By taking advantage of the posterior modes of the latent variables, the initial assumptions of bivariate copula functions are inspected and replaced for better copula functions based on the BIC. Chapter 3 shows an example where the structured factor copula models help to predict the missing temperatures of 24 locations among 479 stations in Germany. The major content of this chapter resulted into a working paper by Nguyen et al. [2018].

Chapter 4 extends the factor copula model to a combination of a factor copula model at the first tree layer and a vine copula structure at a higher tree layer. The model is not only suitable to capture different behaviors at the tail of the distribution but also remains parsimonious with interpretable economic meanings. The truncated factor vine copula models can outperform the multi-factor copula model in cases where there is weak dependence among variables in higher tree levels and the inferences of group latent factors become inaccurate. The VI strategy is used and the dependence structure can be recovered with a similar copula selection procedure. Chapter 4 compares the statistical criteria of different factor models applied to a high dimensional dataset.

Chapter 2

Dynamic one factor copula models

The aim of this chapter is to propose a parallel Bayesian procedure for handling a large set of financial returns using factor copula models. For that, we use EGARCH processes to model the individual returns. Then, the series of standardized innovations are converted into a series of Uniform(0,1) observations, using cumulative distribution functions, that are assumed to have a copula distribution. To handle a large number of returns, we assume a one factor structure that, first, drastically reduces the number of parameters as they scale linearly with the dimension, and, second, provides natural economic interpretations. To account for asymmetric dependence in extreme events, we propose a group dynamic multivariate generalized hyperbolic skew Student-t (MGSt) factor copula where the factor loadings follow Generalized Autoregressive Score (GAS) processes, see Creal et al. [2013] and Harvey [2013]. Importantly, we assume that the dynamic factor loading equation depends on the copula density conditional on the factor rather than the unconditional copula density, as proposed in Oh and Patton [2017b]. The main benefit of our approach is that it allows us to perform parallel inference which greatly reduces the computational cost. Hence, a sizable problem can be fitted from a few minutes up to one hour with a personal computer. The MGSt copula allows for different tail behavior and asymmetric dependence among financial returns.

We compare our proposed dynamic factor copula models with the Exponential Weight Moving

Average (EWMA) and Dynamic Conditional Correlation models (DCC), see Engle and Kelly [2012]. We find that our proposal performs better for high dimensional time series generated in different stress test scenarios. We also consider several copula specifications including the Gaussian and the Student-t as special cases of the generalized hyperbolic Skew Student-t copulas. We show an empirical example of 140 asset returns for companies listed in S&P 500 index. We found the strongest lower tail dependence among stocks in the Insurance and Finance sectors while other sectors such as Food and Beverage, Pharmacy, and Retail only reveal weak lower tail dependence. We also perform optimal portfolio allocation based on minimization of the CVaR. We use the penalized quantile regression method to prevent extreme positive and negative weights. It also overcomes the computational difficulties in comparison with traditional optimization methods.

The rest of the chapter is organized as follows. Section 2.1 introduces the model for univariate marginal returns and specifies our proposal to model the dependence structure with different types of dynamic factor copula models. We present our parallel Bayesian inference strategy in Section 2.2 and describe how to perform return prediction and risk management in Section 2.3. Section 2.4 illustrates the performance of factor copula models with simulated examples. In Section 2.5, we analyze a large series of stock returns listed in S&P 500 and compare the prediction power of models using VaR and CVaR. Section 2.5 also compares the optimal portfolio allocation based on minimizing variance and minimizing CVaR. Finally, conclusions are drawn in Section 2.6.

2.1 Dynamic factor copula models

In this section, we introduce our modeling strategy based on the spirit of Creal and Tsay [2015], Oh and Patton [2017a] and Oh and Patton [2017b]. For that, the first step is to assume a simple $AR - EGARCH$ structure [Nelson, 1991] on the individual returns and then assume a one factor copula structure on the transformed standardized innovations.

2.1.1 Model specification

Let $r_t = (r_{1t}, \ldots, r_{dt})'$, for $t = 1, \ldots, T$, be a d-dimensional financial return time series. We assume that each individual return, r_{it}, for $i = 1, \ldots, d$, follows a stationary $AR(k_i) - EGARCH(p_i, q_i)$ model given by:

$$
\begin{aligned}
r_{it} &= c_i + \phi_{i1} r_{i,t-1} + \ldots + \phi_{ik_i} r_{i,t-k_i} + a_{it} \\
a_{it} &= \sigma_{it} \eta_{it} \\
log(\sigma_{it}^2) &= \omega_i + \sum_{j=1}^{p_i} \beta_{ij} log(\sigma_{i,t-j}^2) + \sum_{j=1}^{q_i} [\alpha_{ij} \eta_{i,t-j} + \gamma_{ij}(|\eta_{i,t-j}| - E|\eta_{i,t-j}|)]
\end{aligned}
$$

where c_i is a constant, $\phi_{i1}, \ldots, \phi_{ik_i}$ are autoregressive parameters verifying the usual stationarity conditions, a_{it} is a sequence of innovations or shocks, σ_{it}^2 is the conditional volatility of the return r_{it}, η_{it} is a sequence of independent standardized innovations with continuous distribution function F_{η_i}, ω_i is a constant, and $\alpha_{i1}, \ldots, \alpha_{iq_i}, \beta_{i1}, \ldots, \beta_{ip_i}, \gamma_{i1}, \ldots, \gamma_{iq_i}$ are $EGARCH$ parameters verifying the usual stationarity conditions. Hence, the $EGARCH$ model takes into account the negative correlation between stock returns and changes in return volatility. We note that the previous $AR - EGARCH$ model can be replaced with any other appropriate specification. For instance, the autoregressive process may be reduced to a simple constant or replaced with an $ARMA$ process, while the $EGARCH$ specification can be replaced with an $GARCH$ [Bollerslev, 1986] or a $GJR - GARCH$ process [Glosten et al., 1993].

Once appropriate models have been specified for all the return series, we can make use of copulas to model their dependence structure. For that, it is well known that $u_{it} = F_{\eta_i}(\eta_{it})$, for each $i = 1, \ldots, d$, is a sequence of independent random variables with a $U(0, 1)$ distribution and the dependence structure among the variables in the vector $u_t = (u_{1t}, \ldots, u_{dt})'$ is given by an unknown copula function. A standard approach is to assume that u_t has either a Gaussian copula or a Student-t copula distribution. Nevertheless, it is questionable whether such copula functions are appropriate. One plausible alternative is to assume, as in Krupskii and Joe [2013],

a factor copula model in which $u_{1t}, \ldots, u_{dt}$ are conditionally independent given a small set of latent variables. Nevertheless, we consider instead an approach in the spirit of Creal and Tsay [2015], Oh and Patton [2017a] and Oh and Patton [2017b]. The idea is to focus on a family of copula models including, among others, the Gaussian, Student-t and generalized hyperbolic skew Student-t copulas, which depend on a conditional scale matrix parameter, R_t, characterized by a factor structure, somehow coming back to standard factor models widely analyzed in the literature. As in Oh and Patton [2017b], we model the dynamic factor loadings as GAS processes, but we assume that the dynamic factor loading equations depend on the copula density conditional on the latent factor rather than the unconditional copula density that allows us to perform parallel inference which heavily reduces the computational cost needed to obtain the conditional posterior distributions of model parameters.

In the next subsections, we describe in detail our proposed dynamic generalized hyperbolic skew Student-t one factor copula model which reduces to Gaussian and Student-t as special cases. We also present some of their advantages over existing alternatives. To simplify, we first present the Gaussian case and then the most general case.

2.1.2 Dynamic Gaussian one factor copula model

In this subsection, we assume that u_t follows a Gaussian copula with correlation matrix parameter R_t and joint distribution function $C(u_{1t}, \ldots, u_{dt} \mid R_t) = \Phi_d\left(\Phi^{-1}(u_{1t}), \ldots, \Phi^{-1}(u_{dt}) \mid R_t\right)$, where $\Phi(\cdot)$ denotes the univariate standard Gaussian cdf and $\Phi_d(\cdot \mid R_t)$ denotes the multivariate Gaussian cdf with zero mean vector and correlation matrix R_t. Therefore, the vector of inverse cdf transformations, $x_t = (x_{1t}, \ldots, x_{dt})'$, where $x_{it} = \Phi^{-1}(u_{it})$, for each $i = 1, \ldots, d$, follows a multivariate Gaussian distribution with zero mean vector and correlation matrix R_t. We assume a dynamic Gaussian one factor copula model for x_t given by:

$$x_t = \rho_t z_t + D_t \epsilon_t, \tag{2.1}$$

where z_t, the latent factor, is a sequence of independent and identically standard Gaussian distributed random variables, $\rho_t = (\rho_{1t}, \ldots, \rho_{dt})'$, is the vector of factor loadings, D_t is a diagonal matrix with elements $\sqrt{1-\rho_{it}^2}$, for $i = 1, \ldots, d$, and $\epsilon_t = (\epsilon_{1t}, \ldots, \epsilon_{dt})'$, is a sequence of independent and identically standard multivariate Gaussian random variables. The latent factor z_t, the idiosyncratic noise ϵ_t, and the dynamic correlations ρ_t are contemporaneously independent. However, the dynamic correlations ρ_t are derived based on the past information of the latent and copula data at time $t-1$. Consequently, the components of the multivariate random vector $x_t = (x_{1t}, \ldots, x_{dt})'$ are conditionally independent given the latent factor z_t and the factor loading vector ρ_t, whose elements, ρ_{it}, represent the correlation between x_{it} and z_t, for $t = 1, \ldots, T$. Therefore, the conditional correlation matrix is given by $R_t = \rho_t \rho_t' + D_t D_t'$. Observe that for the static case, the described model coincides with the one factor Gaussian copula model proposed in Krupskii and Joe [2013]. In a dynamic framework, we allow the components of the correlation vector $\rho_t = (\rho_{1t}, \ldots, \rho_{dt})'$ to vary across time as follows,

$$\begin{aligned} \rho_{it} &= \frac{1 - \exp(-f_{it})}{1 + \exp(-f_{it})} \\ f_{i,t+1} &= (1-b) f_{ic} + a s_{it} + b f_{it} \\ s_{it} &= \frac{\partial \log p(u_t | z_t, f_t, \mathcal{F}_t, \theta)}{\partial f_{it}} \end{aligned} \tag{2.2}$$

for $i = 1, \ldots, d$, where f_{it} is an observation driven process which fluctuates around a constant value f_{ic}, a and b are two parameters that are assumed to be constant across assets, such that $|b| < 1$ to guarantee stationarity, and $p(u_t | z_t, f_t, \mathcal{F}_t, \theta)$ is the conditional probability density function of u_t given the latent variable, z_t, the random vector $f_t = (f_{1t}, \ldots, f_{dt})'$, the set of all information available at time t, denoted by $\mathcal{F}_t = \{U^{t-1}, F^{t-1}\}$, where $U^{t-1} = \{u_1, \ldots, u_{t-1}\}$ and $F^{t-1} = \{f_0, \ldots, f_{t-1}\}$, and the vector of static parameters, $\theta = (a, b, f_{1c}, \ldots, f_{dc})'$. Note that ρ_{it} is assumed to follow a modified logistic transformation, used also in Dias and Embrechts [2010], Patton [2006] and Creal et al. [2013], to guarantee that $\rho_{it} \in (-1, 1)$. Also observe that $f_{i,t+1}$ depends linearly on f_{it} and the adjustment term s_{it}. Clearly, this model reduces to a Gaussian time-invariant one factor

copula model, see Murray et al. [2013] and Oh and Patton [2017a], when $a = b = 0$.

The dynamic equation (2.2) is inspired by the GAS model, see Creal et al. [2013] and Harvey [2013], in which the score s_{it} depends on the complete density of u_t rather than on its first or second moment. Blasques et al. [2015] proved that the use of the score s_{it} leads to the minimum Kullback-Leibler divergence between the true conditional density and the model-implied conditional density, while Koopman et al. [2016] showed some empirical examples where the GAS model outperforms other observation driven models. In addition, we consider here the latent variable z_t as a source of exogenous information and derive the observation density conditional on this source. The main reason for such a choice is to reduce dramatically the computational burden as the score s_{it} has a closed form expression that allows us to parallelize the derivation of the d processes $s_{1t}, \ldots, s_{dt}$. Specifically, as shown in Appendix A.1.1, s_{it} is given by,

$$s_{it} = \frac{1}{2} x_{it} z_t + \frac{1}{2} \rho_{it} - \rho_{it} \frac{x_{it}^2 + z_t^2 - 2\rho_{it} x_{it} z_t}{2\left(1 - \rho_{it}^2\right)}, \tag{2.3}$$

for $i = 1, \ldots, d$. Therefore, s_{it} depends on the values of the pseudo observable x_{it}, the latent variable z_t, and their mutual correlation ρ_{it}. The model is also attractive, as will be shown in the next subsections, s_{it} has a similar structure to the one given in (2.3) for the dynamic Student-t and generalized hyperbolic skew Student-t one factor copula models.

As noted before, the main difference of our proposed model with respect to the dynamic GAS model defined in Oh and Patton [2017b] is that the score in (2.2) is conditioned on the latent variable, z_t. We show in Appendix A.2 that

$$s_{it}^{OP} = \frac{\partial \log p\left(u_t | f_t, \mathcal{F}_t, \theta\right)}{\partial f_{it}} = \mathbf{E}_{z_t}\left[\left.\frac{\partial \log p\left(u_t | z_t, f_t, \mathcal{F}_t, \theta\right)}{\partial f_{it}}\right| u_t, f_t, \theta\right] = \mathbf{E}_{z_t}\left[s_{it} | \; u_t, f_t, \mathcal{F}_t, \theta\right].$$

Thus, the score function (2.3) is the expectation of the proposal score s_{it} over z_t where z_t follows $p(z_t|u_t, f_t, \mathcal{F}_t, \theta)$ distribution. Therefore, since z_t is sampled from its posterior $p(z_t|x_t, f_t, \mathcal{F}_t, \theta)$, one should expect the use of s_{it} in (2.3) to be similar to the use of the score function in Oh and Patton [2017b]. As mentioned before, our proposed specification allows us to obtain the expressions for

s_{it} in parallel for $i = 1, \ldots, d$, reducing the computational burden. This contrasts with Oh and Patton [2017b] where the expressions for s_{it}^{OP} are obtained by the numerical differentiation of the joint copula density, which is much more computationally expensive.

2.1.3 Dynamic generalized hyperbolic skew Student-t one factor copula model

Next, we use the generalized hyperbolic skew Student-t (GSt) distribution proposed by Aas and Haff [2006] to extend the Gaussian factor copula model. The GSt distribution depends on two parameters, ν and γ, which control the generation of extremes events and skewness, respectively. The GSt distribution reduces to the Student-t distribution when $\gamma = 0$ and reduces to the Gaussian distribution when $\gamma = 0$ and $\nu \to \infty$.

Here, we assume that the joint distribution function of u_t is given by $C(u_{1t}, \ldots, u_{dt} \mid R_t, \nu, \gamma) = F_{MGSt}\left(F_{GSt}^{-1}(u_{1t} \mid \nu, \gamma), \ldots, F_{GSt}^{-1}(u_{dt} \mid \nu, \gamma) \mid R_t, \nu, \gamma\right)$, where $F_{GSt}(\cdot \mid \nu, \gamma)$ denotes the univariate standard GSt cdf with degrees of freedom ν and skewness parameter γ, and $F_{MGSt}(\cdot \mid R_t, \nu, \gamma)$ denotes the MGSt cdf with parameters ν and γ and scale matrix R_t. Hence, the MGSt copula allows for asymmetric tail dependence which are not possible with the Gaussian copula assumption. Here, the vector of inverse cdf transformations, $x_t = (x_{1t}, \ldots, x_{dt})'$, where $x_{it} = F_{GSt}^{-1}(u_{it} \mid \nu, \gamma)$, for each $i = 1, \ldots, d$, follows a MGSt with zero location vector, scale matrix R_t, degrees of freedom ν, and skewness parameter γ. Then, we assume a dynamic generalized hyperbolic skew Student-t one factor copula model for x_t given by:

$$x_t = \gamma\zeta_t + \sqrt{\zeta_t}\,(\rho_t z_t + D_t \epsilon_t) \tag{2.4}$$

for $i = 1, \ldots, d$, where z_t, ϵ_t and ρ_t, for $t = 1, \ldots, T$, are as in the Gaussian case, and ζ_t is a sequence of independent and identically inverse Gamma distributed random variables with parameters $\left(\frac{\nu}{2}, \frac{\nu}{2}\right)$, denoted by $IG\left(\frac{\nu}{2}, \frac{\nu}{2}\right)$, and independent of z_t, ϵ_t and ρ_t. Particularly, when $\gamma = 0$, x_t follows multivariate Student-t distribution as a special case. In any case, the components of the multivariate random vector $x_t = (x_{1t}, \ldots, x_{dt})'$ are contemporaneously independent at time t given z_t, ρ_t and ζ_t.

However, note that ρ_t depends on the past values of x_t, z_t and ζ_t through the GAS process.

As in the Gaussian case, the vector $\rho_t = (\rho_{1t}, \ldots, \rho_{dt})'$ is allowed to vary across time as in (2.2), but replacing the value of the score s_{it} with,

$$s_{it} = \frac{\partial \log p\left(u_t | z_t, \zeta_t, f_t, \mathcal{F}_t, \theta\right)}{\partial f_{it}},$$

where $p\left(u_t | z_t, \zeta_t, f_t, \mathcal{F}_t, \theta\right)$ is the conditional probability density function of u_t given z_t, ζ_t, f_t, $\mathcal{F}_t$, and the parameters of the copula function, $\theta = (a, b, \nu, \gamma, f_{1c}, \ldots, f_{dc})'$. Again, this model setting is influenced by the developments in Creal and Tsay [2015] for stochastic factor copulas and Oh and Patton [2017b] for dynamic factor copulas. However, one advantage of our proposal is that the observation driven process remains similar. As shown in Appendix A.1.2, if we let $\tilde{x}_{it} = \frac{x_{it} - \gamma\zeta_t}{\sqrt{\zeta_t}}$, the score function is,

$$s_{it} = \frac{\partial \log p(u_t | z_t, \zeta_t, f_t, \mathcal{F}_t, \theta)}{\partial f_{it}} = \frac{1}{2}\tilde{x}_{it} z_t + \frac{1}{2}\rho_{it} - \rho_{it}\frac{\tilde{x}_{it}^2 + z_t^2 - 2\rho_{it}\tilde{x}_{it} z_t}{2\left(1 - \rho_{it}^2\right)}, \tag{2.5}$$

which is similar to the score function in (2.3). Consequently, we enjoy here the same computational advantages described in the Gaussian case. On the other hand, this proposed model is different from the skew Student-t factor copula model in Oh and Patton [2017a] and Oh and Patton [2017b] since these authors consider different symmetric and asymmetric Student-t distributions for z_t and ϵ_t. Their models do not lead to an easily attainable conditional cdf for x_t and therefore, it is computationally expensive to derive the score s_{it}, as mentioned before.

Demarta and McNeil [2005] noted that the marginal univariate GSt only has finite variance when $\nu > 4$ in comparison with the Student-t distribution which requires $\nu > 2$. They also differ in the tail decay. While the Student-t density has the tail decay as $x^{-\nu-1}$, the GSt density has a heaviest tail decay as $x^{-\nu/2-1}$ and the lightest tail as $x^{-\nu/2-1}\exp\left(-2|\gamma x|\right)$ (for $\gamma \neq 0$). We obtain the tail dependence of the dynamic MGSt one factor copula model using a numerical approximation of the joint quantile exceedance probability, see Appendix A.3. Finally, Demarta and McNeil [2005] suggested several extensions for more complex copula functions. For example, when ζ_t follows

a generalized inverse Gaussian distribution, x_{it} is generalized hyperbolic distributed. Also, one could propose different distributions of the type $x_{it} = \gamma_g h(\zeta_t) + \sqrt{\zeta_t}\left(\rho_{it} z_t + \sqrt{1-\rho_{it}^2}\epsilon_{it}\right)$, where $h(\zeta_t)$ is a function of ζ_t. However, the properties of x_{it} would generally be intractable.

2.1.4 Dynamic group generalized hyperbolic skew Student-t one factor copulas

One potential drawback of the previous models is that only a few parameters control all of the tail co-movements which can be very restrictive for high dimensional returns. In order to relax this assumption, our strategy is to split the d assets into G groups in such a way that returns in the same group have similar characteristics.

Therefore, we write $u_t = (u_{1t}', \ldots, u_{Gt}')'$, where $u_{gt} = \left(u_{1gt}, \ldots, u_{n_g gt}\right)'$, for $g = 1, \ldots, G$ and $\sum_{g=1}^{G} n_g = d$. In the most general case of the MGSt copula, we define $x_{igt} = F_{GSt}^{-1}(u_{igt}|\nu_g, \gamma_g)$ for each asset i, for $i = 1, \ldots, n_g$, belonging to group g, where $g = 1, \ldots, G$, such that,

$$x_{gt} = \gamma_g \zeta_{gt} + \sqrt{\zeta_{gt}}\left(\rho_{gt} z_t + D_{gt}\epsilon_{gt}\right) \tag{2.6}$$

where $x_{gt} = \left(x_{1gt}, \ldots, x_{n_g gt}\right)'$ is the vector of inverse transformations in group g, $\rho_{gt} = \left(\rho_{1gt}, \ldots, \rho_{n_g gt}\right)'$ is the vector of factor loadings in group g, and D_{gt} is the diagonal matrix with elements $\sqrt{1-\rho_{igt}^2}$ and $\epsilon_{gt} = \left(\epsilon_{1gt}, \ldots, \epsilon_{n_g gt}\right)'$ are, respectively, the corresponding diagonal matrix and noise vector in group g.

Observe that the set of mixing variables $\zeta_t = (\zeta_{1t}, \ldots, \zeta_{Gt})'$ create G multivariate MGSt distributions with degrees of freedom parameters $\nu_1, \ldots, \nu_G$ and skewness parameters $\gamma_1, \ldots, \gamma_G$, respectively. Then, the dynamic of the i-th the scale parameters in group g is given by:

$$\begin{aligned} \rho_{igt} &= \frac{1-\exp(-f_{igt})}{1+\exp(-f_{igt})} \\ f_{ig,t+1} &= (1-b_g) f_{igc} + a_g s_{igt} + b_g f_{igt} \end{aligned} \tag{2.7}$$

where the set of parameters $a = (a_1, \ldots, a_G)'$ and $b = (b_1, \ldots, b_G)'$ adjust the dynamic behavior of

the scale parameters in each group g. Here, the i-th score in group g is given by:

$$s_{igt} = \frac{1}{2}\tilde{x}_{igt}z_t + \frac{1}{2}\rho_{igt} - \rho_{igt}\frac{\tilde{x}_{igt}^2 + z_t^2 - 2\rho_{igt}\tilde{x}_{igt}z_t}{2\left(1-\rho_{igt}^2\right)} \tag{2.8}$$

where $\tilde{x}_{igt} = \frac{x_{igt}-\gamma_g\zeta_{gt}}{\sqrt{\zeta_{gt}}}$. Note that when $G = 1$, the model reduces to the copula specification proposed in the previous section.

The model becomes extremely flexible by assuming that each series has its own dynamic group. Indeed, the model is able to capture the different behaviors in the upper and lower tail dependence for those assets in the same group. However, note that the assets in different groups show no tail dependence due to the independence assumption among the components of ζ_t. Also, the pseudo observable $x_{igt} = F_{GSt}^{-1}\left(u_{igt}|\nu_g, \gamma_g\right)$ requires an intensive computation as long as ν_g and γ_g receive new trial values. A parallel Bayesian algorithm is implemented in the next section to speed up calculations.

2.2 Bayesian inference

In this section, we present our parallel Bayesian inference strategy to obtain the posterior distribution of the model parameters of the dynamic one-factor copula models presented in Section 2.1.

2.2.1 Prior distributions

We focus on defining a prior distribution for the copula parameters. In all cases, we use proper but uninformative prior assumptions. We describe the prior for the most general proposed model, the group MGSt factor copula, which contains all other models as particular cases. First, we assume uniform priors for all the elements in $f_c = \{f_{igc} : g = 1, \ldots, G; i = 1, \ldots, n_g\}$. More precisely, we assume a priori that $f_{igc} \sim U\left(-5, 5\right)$, so that the value of ρ_{igc} ranges between $(-0.9866, 0.9866)$. Additionally, f_{11c} is restricted to be positive to guarantee model identifiability. Second, as usual in GAS models, we assume uniform priors for all the elements in $a = \{a_g : g = 1, \ldots, G\}$ and

$b = \{b_g : g = 1, \ldots, G\}$. More precisely, we assume a priori that $a_g \sim U(-0.5, 0.5)$ and $b_g \sim U(0, 1)$. Third, we assume a prior shifted Gamma distributions for all the degrees of freedom parameters in $\nu = \{\nu_g : g = 1, \ldots, G\}$, such that $\nu_g = 4 + \tilde{\nu}_g$, where $\tilde{\nu}_g \sim G(2, 2.5)$, in order that the variance of the pseudo observations, x_{it}, is finite. Fourth, we assume a priori a standard Gaussian distribution for all the skewness parameters in $\gamma = \{\gamma_g : g = 1, \ldots, G\}$, i.e., a priori $\gamma_g \sim N(0, 1)$, for $g = 1, \ldots, G$. In the particular case of a Student-t copula, we assume that ν_g follows a priori a shifted Gamma distribution with $\nu_g = 2 + \tilde{\nu}_g$, such that the variance of x_{it} is finite and set the skewness parameter $\gamma_g = 0$.

Finally, the latent states $z = \{z_t : t = 1, \ldots, T\}$ are treated as nuisance independent parameters following independent $N(0, 1)$ distributions, as considered in the model assumptions. Additionally, the elements of $\zeta = \{\zeta_{gt} : g = 1, \ldots, G; t = 1, \ldots, T\}$ are nested as nuisance parameters for the realization of the pseudo observations x_{it} and depend on the respective elements of ν.

2.2.2 Posterior inference

Given a sample of return data, $r = \{r_t : t = 1, \ldots, T\}$, and the priors defined before, we are interested in the posterior of the model parameters given by the set of marginal parameters, $\vartheta_i = (c_i, \phi_{i1}, \ldots, \phi_{ik_i}, \omega_i, \alpha_{i1}, \ldots, \alpha_{ip_i}, \beta_{i1}, \ldots, \beta_{iq_i}, \gamma_{i1}, \ldots, \gamma_{ip_i})'$, and the set of factor copula parameters, $\vartheta_c = (a, b, \nu, \gamma, z, \zeta, f_c)'$. The likelihood is given by,

$$l(\vartheta_1, \ldots, \vartheta_d, \vartheta_c \mid r) = \prod_{t=1}^{T} c(F_{\eta_1}(\eta_{1t} \mid \vartheta_1), \ldots, F_{\eta_d}(\eta_{dt} \mid \vartheta_d) \mid \vartheta_c) \prod_{i=1}^{d} f_{\eta_i}(\eta_{it} \mid \vartheta_i),$$

where $c(\cdot \mid \vartheta_c)$ denotes the copula density function with parameters ϑ_c and $f_{\eta_i}(\eta_i \mid \vartheta_i)$ is the marginal density function of the standardized innovations, η_{it}. Given this decomposition of the likelihood, we follow the standard two-stage estimation procedure for copulas where, in a first step, we estimate the marginal parameters, $\hat{\vartheta}_i$, independently using the maximum likelihood for each $i = 1, \ldots, d$, and, in a second step, we obtain an approximate sample of the copula observations, $u = \{u_t : t = 1, \ldots, T\}$, where $u_{it} = F_{\eta_i}\left(\eta_{it} \mid \hat{\vartheta}_i\right)$, for $t = 1, \ldots, T$ and for each $i = 1, \ldots, d$. This

two-stage estimation procedure has been shown to be statistically efficient by Joe [2005] and Chen and Fan [2006] in case of parametric and semi-parametric distributions for standardized residuals. Alternatively, a fully Bayesian approach where the joint posterior distribution is approximated in a single step would be done but the two-step approach simplifies enormously the computational burden in the high dimensional setting that we are considering.

Now, considering the G different asset groups, we assume that the matrix sample of copula observations, $u = \{u_t : t = 1, \ldots, T\}$, is such that $u_t = (u'_{1t}, \ldots, u'_{Gt})'$, where $u_{gt} = \left(u_{1gt}, \ldots, u_{n_g gt}\right)'$, for $g = 1, \ldots, G$. Then, the likelihood of the MGSt copula is given by:

$$l\left(\vartheta_c \mid u\right) = \prod_{t=1}^{T} p\left(u_t | z_t, \zeta_t, f_t, \mathcal{F}_t, \theta\right),$$

where $f_t = (f_{1t}, \ldots, f_{Gt})'$ with $f_{gt} = \left(f_{1gt}, \ldots, f_{n_g gt}\right)$, for $g = 1, \ldots, G$. Recall that $\mathcal{F}_t = \{U^{t-1}, F^{t-1}\}$, where $U^{t-1} = \{u_1, \ldots, u_{t-1}\}$ and $F^{t-1} = \{f_0, \ldots, f_{t-1}\}$, and $\theta = (a, b, \nu, \gamma, f_c)'$ is the vector of static parameters. Therefore, given the conditional density (A.1) in Appendix A.1.2, the likelihood is given by:

$$p\left(u | z, \zeta, f_c, a, b, \nu, \gamma\right) = \prod_{t=1}^{T} \prod_{g=1}^{G} \prod_{i=1}^{n_g} \frac{\phi\left(\frac{F_{GSt}^{-1}(u_{igt}|\nu) - \gamma_g \zeta_t}{\sqrt{\zeta_{gt}}} \mid \rho_{igt} z_t, \sqrt{1 - \rho_{igt}^2}\right)}{f_{GSt}\left(F_{GSt}^{-1}\left(u_{igt} \mid \nu_g, \gamma_g\right) \mid \nu_g, \gamma_g\right) \sqrt{\zeta_{gt}}}.$$

As a result, the joint posterior density of the group dynamic MGSt factor copula parameters can be written as follows:

$$\begin{aligned} p\left(z, \zeta, f_c, a, b, \nu, \gamma | u\right) \propto & \prod_{t=1}^{T} \prod_{g=1}^{G} \prod_{i=1}^{n_g} \frac{\phi(\tilde{x}_{igt} | \rho_{igt} z_t, \sqrt{1 - \rho_{igt}^2})}{f_{GSt}\left(x_{igt} | \nu_g, \gamma_g\right) \sqrt{\zeta_{gt}}} \prod_{t=1}^{T} \phi(z_t | 0, 1) \\ & \times \prod_{t=1}^{T} \prod_{g=1}^{G} IG\left(\zeta_{gt} | \frac{\nu_g}{2}, \frac{\nu_g}{2}\right) \prod_{g=1}^{G} G\left(\nu_g - 4 | 2, 2.5\right) \prod_{g=1}^{G} \phi\left(\gamma_g | 0, 1\right), \end{aligned} \tag{2.9}$$

where $\tilde{x}_{igt} = \frac{x_{igt} - \gamma_g \zeta_{gt}}{\sqrt{\zeta_{gt}}}$ and $x_{igt} = F_{GSt}^{-1}\left(u_{it} \mid \nu_g, \gamma_g\right)$.

The computation of the pseudo observable $x_{igt} = F_{GSt}^{-1}(u_{igt} | v_g, \gamma_g)$ is often time-consuming

especially when the value of v_g and γ_g change in each MCMC iteration. We create a sequence of $m = 1000$ values with equal increment in the range $x_{seq} = [x_{Low}, x_{High}]$ and find their exact cdf $u_{seq} = F_{GSt}(x_{seq}|v_g, \gamma_g)$. The approximate values of x_{igt} is calculated as the linear interpolation between two nearest neighbors in the sequence. We employ the algorithm in the SkewHyperbolic package (Scott and Grimson [2015]) to find out the reasonable range $[x_{Low}, x_{High}]$ which guarantees to cover all the values of x_{igt} and also that the relative difference between the approximate and the exact value of x_{igt} is no more than 1%.

2.2.3 MCMC algorithm

Here, a parallel algorithm is exploited to obtain a posterior sample of the model parameters. Due to the fact that the conditional posterior of z_t is Gaussian, we can make fast inference for each latent variable at time $t = 1, \ldots, T$. Also, the conditional posterior of a_g, b_g, ν_g, γ_g, and ζ_{gt} can be sampled in parallel for the groups $g = 1, \ldots, G$, where G is usually a moderate number. Finally, since conditional on z_t, each component of x_t is independent, we can create a parallel estimation procedure for f_{igc} for $i = 1, \ldots, n_g$ and $g = 1, \ldots, G$. Thus, the algorithm is scalable in high dimensional returns.

(i) Set initial values for $\vartheta^{(0)} = \left(z^{(0)}, f_c^{(0)}, a^{(0)}, b^{(0)}, \nu^{(0)}, \gamma^{(0)}, \zeta^{(0)}\right)$.

(ii) For iteration $j = 1, \ldots, N$, obtain $\rho_{igt}^{(j)}$ for $i = 1, \ldots, n_g$, $g = 1, \ldots, G$ and $t = 1, \ldots, T$:

(a) For $t = 1, \ldots, T$, sample $z_t^{(j)} \sim p\left(z_t|u, a^{(j-1)}, b^{(j-1)}, f_c^{(j-1)}, \nu^{(j-1)}, \gamma^{(j-1)}, z_{1:(t-1)}^{(j)}, \zeta^{(j-1)}\right)$.

(b) Parallel for $i = 1, \ldots, n_g$ and $g = 1, \ldots, G$, sample

$$f_{igc}^{(j)} \sim p\left(f_{igc}|u, a^{(j-1)}, b^{(j-1)}, z^{(j)}, \nu^{(j-1)}, \gamma^{(j-1)}, \zeta^{(j-1)}\right).$$

(c) Parallel for $g = 1, \ldots, G$, sample $a_g^{(j)} \sim p\left(a_g|u, b^{(j-1)}, f_c^{(j)}, z^{(j)}, \nu^{(j-1)}, \gamma^{(j-1)}, \zeta^{(j-1)}\right)$.

(d) Parallel for $g = 1, \ldots, G$, sample $b_g^{(j)} \sim p\left(b_g|u, a^{(j)}, f_c^{(j)}, z^{(j)}, \nu^{(j-1)}, \gamma^{(j-1)}, \zeta^{(j-1)}\right)$.

(e) Parallel for $g = 1, \ldots, G$, sample $\nu_g^{(j)} \sim p\left(\nu_g | u, a^{(j)}, b^{(j)}, f_c^{(j)}, z^{(j)}, \gamma^{(j-1)}, \zeta^{(j-1)}\right)$.

(f) Parallel for $g = 1, \ldots, G$, sample $\gamma_g^{(j)} \sim p\left(\gamma_g | u, a^{(j)}, b^{(j)}, f_c^{(j)}, z^{(j)}, \nu^{(j)}, \zeta^{(j-1)}\right)$.

(g) Parallel for $g = 1, \ldots, G$, sample $\zeta_{gt}^{(j)} \sim p\left(\zeta_{gt} | u, a^{(j)}, b^{(j)}, f_c^{(j)}, z^{(j)}, \nu^{(j)}, \gamma^{(j)}, \zeta_{g,1:(t-1)}^{(j)}\right)$ for $t = 1, \ldots, T$.

The conditional posterior distributions for all the parameters are given in Appendix A.4. In the algorithm, we apply the Gibbs sampler for step $2a$ and the Adaptive Random Walk Metropolis Hasting (ARWMH) (see Roberts and Rosenthal [2009]) for steps $2b$ to $2f$. As suggested by Creal and Tsay [2015], we use the independent MH in step $2g$ to generate new values of $\log\left(\zeta_{gt}^{(j)}\right)$ from a Student-t distribution with 4 degrees of freedom with mean equal to the mode and scale equal to the inverse Hessian at the mode. Logarithms guarantee that $\zeta_{gt}^{(j)}$ is positive. Thus, for each time period t, we accept $\zeta_{gt}^{(j)}$ with probability:

$$\min\left\{1, \frac{p\left(\zeta_{gt}^{(j)} | u, a^{(j)}, b^{(j)}, f_c^{(j)}, z^{(j)}, \nu^{(j)}, \gamma^{(j)}, \zeta_{g,1:(t-1)}^{(j)}\right) q\left(\zeta_{gt}^{(j-1)}\right)}{p\left(\zeta_{gt}^{(j-1)} | u, a^{(j)}, b^{(j)}, f_c^{(j)}, z^{(j)}, \nu^{(j)}, \gamma^{(j)}, \zeta_{g,1:(t-1)}^{(j)}\right) q\left(\zeta_{gt}^{(j)}\right)}\right\}.$$

Observe that this Bayesian algorithm reduces to steps $2a$ to $2d$ for the dynamic Gaussian one factor copula. Also, step $2f$ is omitted for the dynamic Student-t one factor copula since $\gamma = 0$. The codes and implementation of the algorithm are available at https://github.com/hoanguc3m/FactorCopula.

2.3 Prediction of returns and risk management

In this section, we illustrate how the estimated copula models help to predict returns and measure the risk of the portfolio such as portfolio variance, quantile of the portfolio's profit/loss distribution for a given horizon (VaR) and conditional expected loss above a quantile (CVaR). Finally, we employ a simulation procedure to allocate an optimal portfolio based on minimum variance and minimum CVaR.

2.3.1 Prediction of returns

Based on the MCMC samples from the conditional posterior distribution of copula parameters $\vartheta_c^{(n)} = \left(a^{(n)}, b^{(n)}, \nu^{(n)}, \gamma^{(n)}, z^{(n)}, \zeta^{(n)}, f_c^{(n)}\right)$, for $n = 1, \ldots, N$, we can obtain the distribution of the predicted return $r_t = \{r_{i,t} : i = 1, \ldots, d\}$ at time $t = T + 1$. For the sake of simplicity, we consider $AR(1) - EGARCH(1,1)$ for the marginal and generate replications of one-step-ahead predicted return $(r_{1t}^{(n)}, \ldots, r_{dt}^{(n)})$ as follows,

$$
\begin{aligned}
r_{it}^{(n)} &= \hat{c}_i + \hat{\phi}_{i1} r_{i,t-1} + a_{it}^{(n)} \\
a_{it}^{(n)} &= \sigma_{it} \eta_{it}^{(n)} \\
log(\sigma_{it}^2) &= \hat{\omega}_i + \hat{\alpha}_{i1} \eta_{i,t-1} + \hat{\gamma}_{i1}(|\eta_{i,t-1}| - E|\eta_{i,t-1}|) + \hat{\beta}_{i1} log(\sigma_{i,t-1}^2)
\end{aligned}
$$

where $\hat{\vartheta}_i = \left(\hat{c}_i, \hat{\phi}_{i1}, \hat{\omega}_i, \hat{\alpha}_{i1}, \hat{\beta}_{i1}, \hat{\gamma}_{i1}\right)'$ is the set of marginal parameters in $AR(1) - EGARCH(1,1)$ model. The standardized innovation is obtained as $\eta_{it}^{(n)} = F_{\eta_i}^{-1}(u_{igt}^{(n)}) = F_{\eta_i}^{-1}(F_{GSt}(x_{igt}^{(n)}|\vartheta_c^{(n)}))$ and the value of $x_{igt}^{(n)}$ is generated from Equations (2.6 - 2.8) where $\zeta_{gt}^{(n)} \sim IG(\frac{\nu^{(n)}}{2}, \frac{\nu^{(n)}}{2})$, $z_t^{(n)} \sim N(0,1)$, and $\epsilon_{igt}^{(n)} \sim N(0,1)$, i.e.,

$$
\begin{aligned}
x_{igt}^{(n)} &= \gamma_g^{(n)} \zeta_{gt}^{(n)} + \sqrt{\zeta_{gt}^{(n)}} \left(\rho_{igt}^{(n)} z_t^{(n)} + \sqrt{(1 - \rho_{igt}^{(n)2})} \epsilon_{igt}^{(n)}\right) \\
\rho_{igt}^{(n)} &= \frac{1 - \exp\left(-f_{igt}^{(n)}\right)}{1 + \exp\left(-f_{igt}^{(n)}\right)} \\
f_{igt}^{(n)} &= \left(1 - b_g^{(n)}\right) f_{igc}^{(n)} + a_g^{(n)} s_{ig,t-1}^{(n)} + b_g^{(n)} f_{ig,t-1}^{(n)} \\
s_{ig,t-1}^{(n)} &= \frac{1}{2} \tilde{x}_{ig,t-1}^{(n)} z_{t-1}^{(n)} + \frac{1}{2} \rho_{ig,t-1}^{(n)} - \rho_{ig,t-1}^{(n)} \frac{\tilde{x}_{ig,t-1}^{(n)2} + z_{t-1}^{(n)2} - 2\rho_{ig,t-1}^{(n)} \tilde{x}_{ig,t-1}^{(n)} z_{t-1}^{(n)}}{2\left(1 - \rho_{ig,t-1}^{(n)2}\right)} \\
\tilde{x}_{ig,t-1}^{(n)} &= \frac{x_{ig,t-1}^{(n)} - \gamma_g^{(n)} \zeta_{g,t-1}^{(n)}}{\sqrt{\zeta_{g,t-1}^{(n)}}} = \frac{F^{-1}(u_{ig,t-1}|\gamma_g^{(n)}, \nu_g^{(n)}) - \gamma_g^{(n)} \zeta_{g,t-1}^{(n)}}{\sqrt{\zeta_{g,t-1}^{(n)}}}
\end{aligned}
$$

We can also obtain the distribution of predicted return at time $T + h$, where $h > 1$, conditional on the return information at time $t = T + h - 1$. As the return prediction needs information about

the latent variables z_t and ζ_{gt}, we choose z_t and ζ_{gt} as the maximum a posteriori of its conditional posterior distribution when obtaining new data.

2.3.2 Risk measurement

Assume that we have a portfolio constructed with the return series $r_{1t}, \ldots, r_{dt}$. Then, the total return at time t is calculated as,

$$r_t = \sum_{i=1}^{d} \delta_{it} r_{it}$$

where $\delta_t = \{\delta_{it}\}_{i=1}^{d}$ is the set of asset weights in the portfolio at time t such that $\sum_{i=1}^{d} \delta_{it} = 1$. The $q\%$ VaR is the threshold loss value such that the probability of a loss exceeds VaR is q, over the time horizon t, i.e.,

$$q = \texttt{Pr}\left(\sum_{i=1}^{d} \delta_{it} r_{it} \leq -\texttt{VaR}_{q,t}\right).$$

Similarly, the CVaR is the conditional expected loss above $q\%$ VaR, i.e.,

$$\texttt{CVaR}_{q,t} = -\texttt{E}\left(\sum_{i=1}^{d} \delta_{it} r_{it} \middle| \sum_{i=1}^{d} \delta_{it} r_{it} \leq -\texttt{VaR}_{q,t}\right).$$

Here, we estimate the one-step-ahead $VaR_{q,t}$ and $CVaR_{q,t}$ for the portfolio of equal weight. In the previous section, we obtain the distribution of one-step-ahead predicted return $\{(r_{1,t}^{(n)}, \ldots, r_{d,t}^{(n)})\}_{t=T+1}^{T+H}$. Then, it is easy to obtain the predictive $\text{VaR}_{q,t}$ and $\text{CVaR}_{q,t}$ using the return simulation. The estimated VaR_q and CVaR_q are the average of $\{\text{VaR}_{q,t}\}_{t=T+1}^{T+H}$ and $\{\text{CVaR}_{q,t}\}_{t=T+1}^{T+H}$ along the prediction period. We compare the prediction powers of the proposed copula models using backtesting for VaR. The expected number of days that the realized portfolio return goes below the $VaR_{q,t}$ threshold is qH.

2.3.3 Optimal portfolio allocation

Next, we can take advantage of the predicted returns above for active portfolio allocation. Classically, Markowitz [1952] introduces portfolio allocation theory based on the mean-variance approach.

The optimal weight for the minimum mean-variance problem is obtained by solving

$$\hat{\delta}_t = \arg\min_{\delta_t} \left\{ \delta_t' \Sigma_t \delta_t : \delta_t' \mathbf{1} = 1, \delta_t' \mu_t = \mu_0 \right\}$$

where μ_t and Σ_t are the expected return and the covariance matrix of the assets in the portfolio at time t, and μ_0 is the expected return. Jagannathan and Ma [2003] recommend imposing nonnegative constraints on portfolio weights ($\delta_t > 0$). This strategy is not only commonly used by practitioners but also improves the efficiency of optimal portfolios using sample moments. In the empirical illustration, we show an example of an optimal portfolio using minimum variance, as follows:

$$\hat{\delta}_t^{(Var)} = \arg\min_{\delta_t} \left\{ V \left(\sum_{i=1}^{d} \delta_{it} r_{it} \right) : \delta_t' \mathbf{1} = 1, \delta_t' \geq 0 \right\}$$

Alternatively, the common optimization problem is to obtain the portfolio with minimum VaR or CVaR. Alexander and Baptista [2004] compare the portfolio selection using VaR and CVaR and recommend CVaR as a more appropriate tool for risk management. However, the minimum CVaR portfolio is often time consuming in high dimensions and results in extreme asset weights. Xu et al. [2016] deal with this issue by proposing a weight constraint on the minimum CVaR portfolio,

$$\hat{\delta}_t^{(CVaR)} = \arg\min_{\delta_t} \left\{ \texttt{CVaR}_{q,t} + \lambda_t \sum_{i=1}^{d} Pen(\delta_{it}) : \delta_t' \mathbf{1} = 1 \right\}$$

where λ_t is a penalty parameter and the Pen function can be chosen as the LASSO (Tibshirani [1996]), or SCAD (Fan and Li [2001]) penalty functions, among others. Following Bassett Jr et al. [2004], the CVaR can be written as,

$$\texttt{CVaR}_{q,t} = q^{-1} \arg\min_{\xi_t} E\rho_q \left[r_t - \xi_t \right] - \mu_t$$

where ξ_t is the q quantile of r_t. The quantile loss function $\rho_q [u] = u(q - I(q < 0))$ as defined in

Koenker and Bassett Jr [1978]. Note that

$$r_t = \sum_{i=1}^{d} \delta_{it} r_{it} = r_{1t} - \sum_{i=2}^{d} \delta_{it}(r_{1t} - r_{it})$$

Let $Y_t = r_{1t}$ and $X_{it} = r_{1t} - r_{it}$. Then, it is straightforward to write the optimal portfolio problem with LASSO penalty as a Lasso penalized quantile regression,

$$\hat{\delta}_t^{(CVaR)} = \arg\min_{\delta_t, \xi_t} E\rho_q \left[Y_t - \sum_{i=2}^{d} \delta_{it} X_{it} - \xi_t \right] - \lambda_t \sum_{i=1}^{d} |\delta_{it}|$$

where the factor q is absorbed into the penalty term, and μ_t is the constant at time t. We choose a λ_t for each period based on the minimum BIC value for the penalized quantile regression (see Lee et al. [2014])

$$\hat{\lambda}_t = \arg\min_{\lambda_t} \log \left(\sum_{n=1}^{N} \rho_\tau (Y_t^{(n)} - \sum_{i=2}^{d} \delta_{it} X_{it}^{(n)} - \xi_t) \right) + |S| \frac{\log N}{2N}$$

where N is the number of return simulation and $|S|$ is the number of points in the set S such that $S = \{i : \hat{\delta}_{it,\lambda} \neq 0, i \in [2, p]\}$. We substitute the optimal weights in each period to obtain $CVaR_{q,t}$

2.4 Simulation study

2.4.1 Simulated data

In this section, we illustrate the proposed Bayesian methodology using simulated data from the MGSt one factor copula in Section 2.1.4. We generate a random sample of $d = 100$ time series with $G = 10$ groups of different sizes and a time length $T = 1000$ from Equations (2.6) to (2.8). The value of the parameters $\vartheta_c = (a, b, \nu, \gamma, z, \zeta, f_c)'$ are randomized. More precisely, a is generated from a $U(0.05, 0.10)$ distribution, b is generated from a $U(0.95, 0.985)$, ν is generated from a $U(6, 18)$, γ is generated from a $U(-1, 0)$ distribution, z_t is generated from a $\Phi(0, 1)$ distribution, and ζ_{gt} is generated from an $IG(\nu_g/2, \nu_g/2)$ where $t = 1, \ldots, T, g = 1, \ldots, G$. The expected correlation between pseudo observation x_t and the latent factor z_t are sampled from a $U(0.1, 0.9)$ distribution,

which results in values for f_{igc} ranging in the interval $(0.2, 3)$.

We estimate the set of true parameters, ϑ_c, using 20.000 MCMC iterations where the first 10.000 are discarded as burn-in iterations. The algorithm seems to perform adequately and convergence is fast. Practically, all the posteriors reached convergence after 1000 iterations. We retain every 10-th iterations to reduce autocorrelation. The algorithm takes around 25 minutes, 70 minutes and 90 minutes for the Gaussian, Student and MGSt one factor copula model, respectively, on an Intel Core i7-4770 processor (4 cores - 8 threads - 3.4GHz).

Figure 2.1 shows the box plots of the posterior sample from the MCMC together with the true values of the model parameters. Observe that the true values of a_g, b_g, ν_g and γ_g lie between the first and the third quantile of the credible intervals in 50% of the cases and never reach out of their whiskers. The posterior distributions of b_g are skewed to the left with heavier tails. Also, the posterior samples show larger variances for higher values of the degree of freedom parameters ν_g. We have observed that there is a negative correlation between MCMC samples of ν_g and γ_g which means that if the posterior mean of ν_g underestimates its true value, the value of γ_g will overestimate its true value. However, the effect is weakly observed. We select some values of f_c and z_t to illustrate the comparison between the posterior mean of f_{igc} versus its true value, for $i = 1, \ldots, d$ and $g = 1 \ldots, G$, and z_t versus its true value, for $t = 1, \ldots, T$. We obtain quite accurate results. The posterior variance of z_t also reduces when the dimension increases. We obtain a smaller posterior standard deviation of ρ_c when its true value is high. In general, most of the parameters which govern the dynamic dependence in each group are correctly estimated. We perform a Monte Carlo study in Online Appendix.

2.4.2 Comparison of estimators

Next, we compare several dynamic correlation models in different scenarios based on Engle [2002] and Creal et al. [2011]'s proposal. We generate $d = 10$ time series from a multivariate Gaussian,

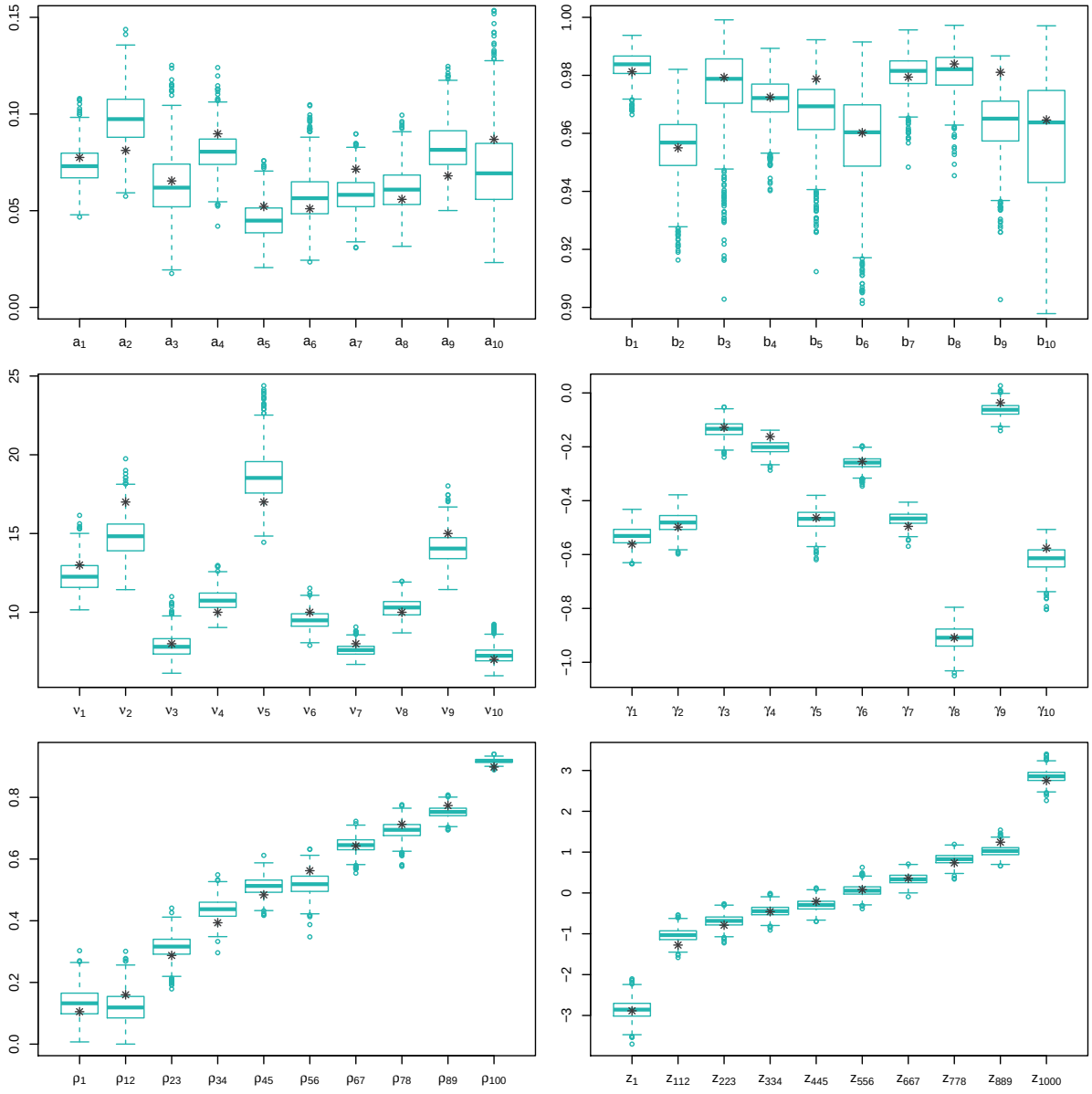

Figure 2.1: Box plots for the posterior samples of $(a, b, \nu, \gamma, \rho_c, z)$ and true values (stars)
The figure shows the box plot for the posterior samples from simulated data. We select a few parameters ρ_c and z based on their ranks of values to conserve space. We observe that the true values of (a, b, ν, γ) never reach out of their whiskers. The online version of this figure is in color.

Student-t, GSt distribution with zero location vector and scale matrix R_t such that,

$$\text{Multivariate Gaussian: } y_t = R_t^{1/2}\epsilon_t$$

$$\text{Multivariate Student-}t : y_t = \sqrt{\zeta_t}R_t^{1/2}\epsilon_t$$

$$\text{Multivariate GSt: } y_t = \gamma\zeta_t + \sqrt{\zeta_t}R_t^{1/2}\epsilon_t$$

where $\epsilon_t \sim \Phi_d(0, I_d)$, $\zeta_t \sim IG(\nu/2, \nu/2)$, we let $\nu = 10$ and $\gamma = -0.1$ and the time varying scale matrix

$$R_t = \begin{bmatrix} 1 & ... & r_{1dt} \\ ... & 1 & ... \\ r_{d1t} & ... & 1 \end{bmatrix} = \rho_t \rho_t' + D_t D_t' \tag{2.10}$$

where D_t is a diagonal matrix with elements $1 - \rho_t^2$. We consider six models to account for different behaviors of $\rho_t = (\rho_{1t}, \ldots, \rho_{dt})'$ in high dimensions, for $\forall i = 1, \ldots, d$:

(i) Constant: $\rho_{it} = \sqrt{0.9}$.

(ii) Sine: $\rho_{it} = \sqrt{0.5 + 0.4cos(2\pi t/200 + \varphi i/20)}$.

(iii) Fast sine: $\rho_{it} = \sqrt{0.5 + 0.4cos(2\pi t/20 + \varphi i/20)}$.

(iv) Step: $\rho_{it} = \sqrt{0.4 + 0.5I(t > 500)}$.

(v) Ramp: $\rho_{it} = \sqrt{mod((t + \varphi i)/200)}$.

(vi) Model: $\rho_{it} = \sqrt{\frac{\exp(h_{it})}{1+\exp(h_{it})}}$ where $h_{it} = -0.4(1 - 0.99) + 0.99h_{i,t-1} + 0.14\eta_{it}$ and $\eta_{it} \sim \Phi(0, 1)$.

Here, φ is used to control for the co-movement of the joint dependence of time series. When $\varphi = 0$, we have equivalent-scale models in sine, fast sine, and ramp. Figure 2.2 shows the r_{ijt} process for selected elements in the scale matrix R_t with $\varphi = 10$.

We generate 100 datasets for each multivariate distribution and estimate the scale matrix R_t using the EWMA, the DCC (Engle [2002]) and the GAS models. For EWMA, we let

$$\Sigma_t = \phi\Sigma_{t-1} + (1 - \phi)y_{t-1}y_{t-1}' \text{ where } \phi = 0.96,$$
$$R_t = diag(\Sigma_t)^{-1/2}\Sigma_t diag(\Sigma_t)^{-1/2}.$$

We generate 1100 time periods and measure the accuracy of each model based on the mean absolute

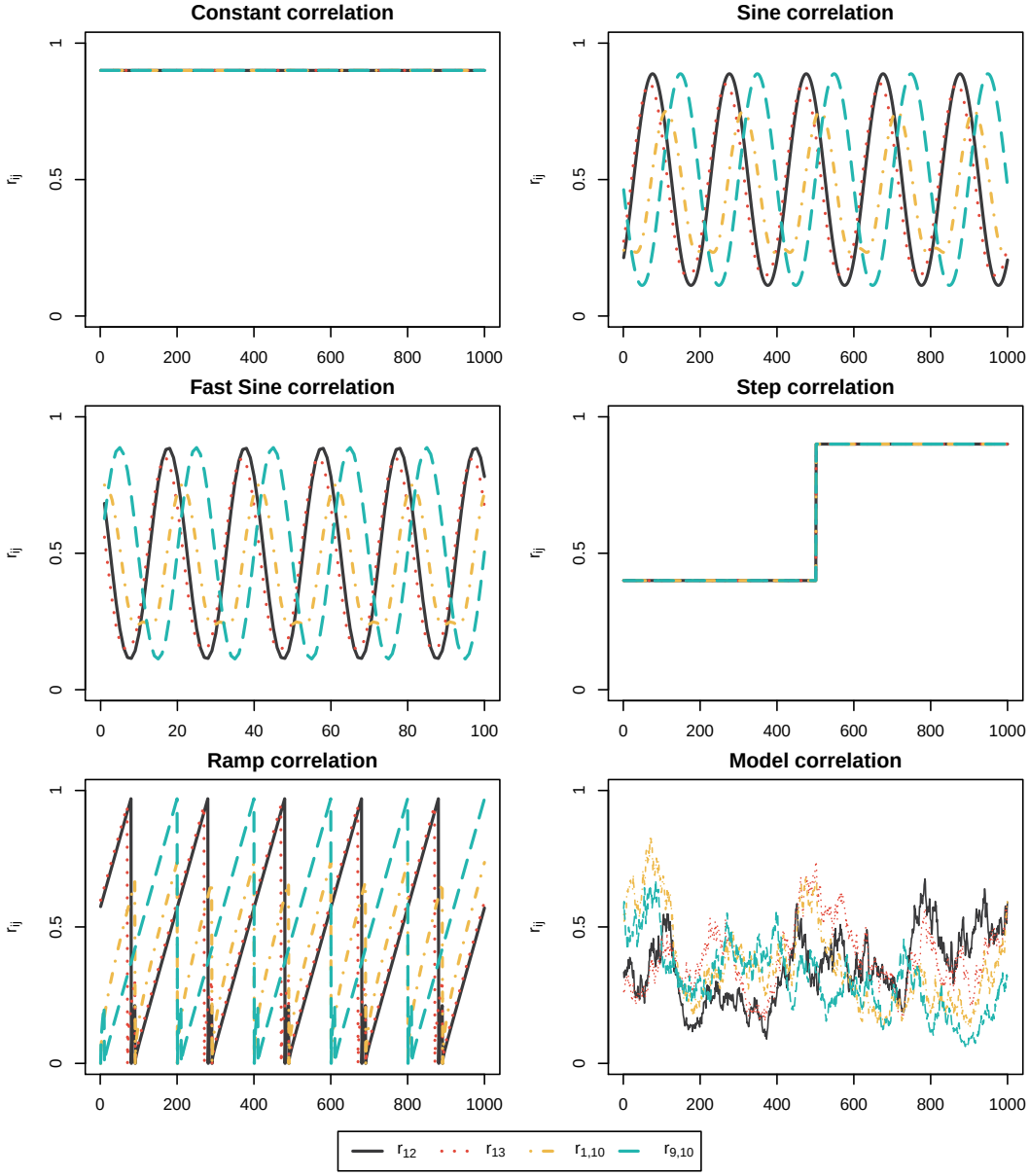

Figure 2.2: The r_{ij} processes for different stress tests
The figure shows selected r_{ij} processes for different stress scenarios. The R_t is equivalent-scale matrix in case of constant and step correlations. For sine, fast sine and ramp models, we set $\varphi = 10$ to account for lag in the correlation. The online version of this figure is in color.

error (MAE) and the mean squared error (MSE) for the last $T = 1000$ observations,

$$MAE = \frac{1}{T} \frac{\sum_{t=1}^{T} \sum_{i=1}^{d} \sum_{j=1}^{d} (|\hat{\rho}_{ijt} - \rho_{ijt}|)}{d^2 - d},$$

$$MSE = \frac{1}{T} \frac{\sum_{t=1}^{T} \sum_{i=1}^{d} \sum_{j=1}^{d} (\hat{\rho}_{ijt} - \rho_{ijt})^2}{d^2 - d}.$$

Table 2.1 shows the comparison among the GAS and the EWMA to the benchmark DCC model. The relative values of MAE and MSE are measured based on the mean of relative MAE and MSE for each Monte Carlo dataset. For multivariate Gaussian simulation, the GAS model is preferred over the DCC model with an increase of accuracy at least 10% with $\varphi = 0$. The GAS model also outperforms the DCC model in most of the scenarios with multivariate Student-t and MGSt distribution for $\varphi = 10$.

Table 2.1: MAE and MSE results: in-sample

	Constant	Sine	Fast sine	Step	Ramp	Model
MAE - Multivariate Gaussian						
GAS	0.562	0.639	0.794	0.765	0.738	0.861
DCC	1.000	1.000	1.000	1.000	1.000	1.000
EWMA	5.153	1.048	1.068	1.068	0.921	1.199
MSE - Multivariate Gaussian						
GAS	0.384	0.563	0.721	0.646	0.809	0.829
DCC	1.000	1.000	1.000	1.000	1.000	1.000
EWMA	31.367	1.173	1.210	1.080	1.090	1.484
MAE - Multivariate Student-t						
GAS	0.506	0.792	0.883	0.871	0.805	0.947
DCC	1.000	1.000	1.000	1.000	1.000	1.000
EWMA	4.636	1.160	1.193	1.043	1.059	1.251
MSE - Multivariate Student-t						
GAS	0.321	0.738	0.843	0.793	0.876	0.985
DCC	1.000	1.000	1.000	1.000	1.000	1.000
EWMA	27.077	1.395	1.495	1.109	1.271	1.633
MAE - MGSt						
GAS	0.867	0.878	0.981	0.953	0.892	0.916
DCC	1.000	1.000	1.000	1.000	1.000	1.000
EWMA	5.135	1.107	1.194	1.137	1.045	1.181
MSE - MGSt						
GAS	0.483	0.869	0.936	0.943	0.964	0.905
DCC	1.000	1.000	1.000	1.000	1.000	1.000
EWMA	31.248	1.322	1.495	1.323	1.252	1.371

The table shows the MAE and MSE for the estimated dynamic scale matrix of EWMA, DCC, GAS models. The MAE and MSE are measured as relative values with respect to DCC model. For sine, fast sine, ramp model, we set $\varphi = 0$ for multivariate Gaussian distribution and $\varphi = 10$ for multivariate Student-t and multivariate GSt distribution.

2.5 Empirical data

In this section, we illustrate our approach with a series of $d = 140$ daily stock returns of companies listed in the S&P 500 index, from 01/01/2007 to 01/09/2014. This period includes the subprime mortgage crisis (2007-2009) and the European sovereign debt crisis (2010-2012). The data are taken from Datastream and contain $T = 2000$ days observed during the considered eight-year period. Table 2.2 shows the summary statistics for the daily stock returns. The mean daily return over the 140 stocks is 0.056%. The most extreme individual events are a one day crash of -53.8% and a one day gain of 87.0%. We calculate the robust skewness and robust kurtosis based on the quantile distribution of returns due to outliers, see Kim and White [2004]. The average skewness is 0.05, which reflects a slight asymmetry of the observed returns, and the average excess kurtosis is 0.312, which shows the heavy tails of the return distributions.

As described in Section 2.2, we use a two-stage procedure to estimate the dependence structure of the stock returns. In Subsection 2.5.1, we fit a $AR\,(1)-EGARCH\,(1,1)$ model for the conditional mean and variance of marginal returns. Then, we take out the standardized innovations and transform them into the copula observations using the corresponding cdf. In Subsection 2.5.2, we estimate the one factor copula models and illustrate some empirical findings.

Table 2.2: Summary statistics for cross-sectional daily returns of 140 firms listed in S&P 500

Statistics	N	Mean	Minimum	1st Qu.	Median	3rd Qu.	Maximum
Mean	140	0.056	-0.028	0.038	0.053	0.071	0.168
Minimum	140	-16.559	-53.802	-19.929	-14.845	-11.470	-5.372
1. Quartile	140	-0.928	-1.801	-1.094	-0.890	-0.708	-0.417
Median	140	0.009	-0.000	0.000	0.000	0.011	0.066
3. Quartile	140	1.041	0.511	0.837	1.011	1.223	1.852
Maximum	140	19.435	7.795	13.015	17.941	22.854	86.983
Skewness	140	0.050	-0.045	0.035	0.054	0.071	0.131
Excess Kurtosis	140	0.312	0.124	0.256	0.303	0.359	0.661

Summary statistics for cross-sectional daily returns (in percentages) of 140 firms listed in S&P 500 index. The distribution of the common statistics are shown by rows label using quantiles and mean. The outliers can distort the sample statistics, hence we calculate the robust skewness (RS_2) and robust kurtosis (RK_2) based on the quantile distribution of returns, see Kim and White [2004]

2.5.1 Marginal distributions

For simplicity, we fit an $AR\,(1) - EGARCH\,(1,1)$ - skew Student-t model for each marginal return series using $rugarch$ package, see Ghalanos [2018]. The standardized innovation, η_{it}, is assumed to follow a univariate skew Student-t distributions (sStd), see Fernández and Steel [1998], with degrees of freedom $\nu_{i\eta}$ and skewness parameter $\xi_{i\eta}$, for $i = 1, \ldots, d$. Then,

$$\begin{aligned} r_{it} &= c_i + \phi_{i1} r_{i,t-1} + \sigma_{it}\eta_{it} \\ log(\sigma_{it}^2) &= \omega_i + \alpha_{i1}\eta_{i,t-1} + \gamma_{i1}(|\eta_{i,t-1}| - E|\eta_{i,t-1}|) + \beta_{i1} log(\sigma_{i,t-1}^2) \\ \eta_{it} &\sim skew - Student - t(\nu_{i\eta}, \xi_{i\eta}) \\ u_{it} &= F_{sStd}(\eta_{it}|\nu_{i\eta}, \xi_{i\eta}) \end{aligned}$$

Table 2.3 shows the summary statistics for the posterior mean estimations of the univariate $AR\,(1) - EGARCH\,(1,1)$ - skew Student-t model across the 140 firms. The effect of the conditional autoregressive mean is weak as the average value of ϕ_{i1} is -0.035 and ϕ_{i1} is insignificantly different from 0 in most of marginals. This fact matches with other findings in the finance literature that the return levels are unpredictable. However, the variance returns are quite predictable through the $EGARCH\,(1,1)$ setting. The average of α_{i1}, β_{i1}, and γ_{i1} are significantly different from 0 for most of the marginals. On average, the volatility clustering is captured by the parameter β_{i1} standing at 0.989. The degrees of freedom for each marginal also diversifies between 2.66 and 9.86 which accounts for different kurtosis. The skewness parameter ranges from 0.871 to 1.109 and most of them are not significantly different from the one which represents for symmetric Student-t distribution. However, the leverage effect that negative return usually leads to an increase in the volatility of innovation is significantly observed for all marginal returns.

For the second stage, the standardized innovations are taken out and transformed to copula observations by applying the corresponding marginal cdfs. More specifically, using the maximum likelihood estimations for each marginal, we obtain the standardized innovations, η_{it}, of the $AR\,(1) - EGARCH\,(1,1)$ process and transform them into the copula observations $u_{it} = F_{sStd}\left(\eta_{it}|\hat{\vartheta}_i\right)$

Table 2.3: Summary statistics of $AR\,(1) - EGARCH\,(1,1)$-skew Student-t for marginal returns

	Mean	Minimum	1. Quartile	Median	3. Quartile	Maximum
c_i	0.054	-0.028	0.033	0.050	0.067	0.170
ϕ_{i1}	-0.035	-0.127	-0.056	-0.037	-0.012	0.053
ω_i	0.011	-0.011	0.007	0.009	0.013	0.062
α_{i1}	-0.066	-0.131	-0.081	-0.067	-0.052	-0.022
β_{i1}	0.989	0.959	0.986	0.991	0.994	0.999
γ_{i1}	0.137	0.036	0.113	0.135	0.163	0.246
$\xi_{i\eta}$	0.998	0.871	0.972	0.995	1.022	1.109
$\nu_{i\eta}$	5.407	2.660	4.479	5.159	5.926	9.858

Summary statistics for the maximum likelihood estimations of the univariate $AR\,(1) - EGARCH\,(1,1)$-skew Student model across the 140 firms. The distribution of estimations are described by mean and quantiles.

where $\hat{\vartheta}_i = \{\hat{c}_i, \hat{\phi}_{i1}, \hat{\omega}_i, \hat{\alpha}_{i1}, \hat{\beta}_{i1}, \hat{\gamma}_{i1}, \hat{\nu}_{i\eta}, \hat{\xi}_{i\eta}\}$. This simplifies the computational burden in the high dimensional setting in which we only concentrate on estimating the copula parameters. Apart from that, we check if the choice of the sStd distribution is suitable with univariate GARCH volatilities by performing the Kolmogorov-Smirnov goodness of fit test as well as the Anderson-Darling test, the Neyman's smooth test of fit in Online Appendix. All series passed the test with p-values larger than 0.05. We also tested for the serial correlation of the innovations and did not find significant results.

2.5.2 Copula estimation

Next, we apply the proposed Bayesian approach to nine different one factor copula models. These are the dynamic Gaussian, Student-t and MGSt combined with particular cases of models referred as block equivalent mean correlation, single group and multiple-group. In the block equi-mean correlation model, the parameter f_{igc} in (2.7) is restricted to be the same for the assets belonging to the same group, i.e., $f_{1gc} = \ldots = f_{n_g gc}$, for all $g = 1, \ldots, G$. We classified $G = 12$ group industries of assets depending on their SIC codes, as in Creal and Tsay [2015], Oh and Patton [2017b], among others. These are Oil & Construction , Food & Beverage, Pharmaceuticals, Plastic Material & Plant Chemical, Textile & Papers, Steel, Home Appliances & Automobile, Electronics, Transportation

& Communication, Retail & Distribution, Insurance, Finance (not contained in Insurance). The detailed number of firms are reported in Online Appendix. On average, there are 12 firms in each sector group. In the single group model, $G = 1$ as described in (2.4), only a few numbers of parameters account for the tail dependence, while the correlations are allowed to be different across the assets. Finally, the multiple-group dynamic model is the most flexible one with different behaviors in the tail dependence and unrestricted scale parameters as in (2.7). In all cases, we generate $40,000$ iterations with $10,000$ burn-in for each factor copula model and every 10-th draws are taken in order to prevent the autocorrelation in the MCMC chains. We also check for the convergence of the MCMC chains in Online Appendix.

Table 2.4 outlines the main estimation results for the nine one factor copula models. In particular, the table includes the value of the AIC, BIC, and DIC for the model selection, obtained as explained in Appendix A.5. The dynamic MGSt copula appears to show a better fit over the Gaussian and Student-t copula models. In general, the posterior means of the parameters a, b and f_c are similar across Gaussian, Student-t, and MGSt copulas, as shown for example in the model (3), (6) and (9). The block equi-mean correlation model reports a smaller posterior range for a and b. The degrees of freedom and skewness parameters are roughly similar between the block-equivalent and the multiple-group models. The model selection criteria show an interesting result that the models with more parameters accounting for extreme events are preferable over the models that have limitations on these behaviors. For example, in Gaussian copulas, the one group outperforms the block-equivalent due to the fact that they do not capture the extreme occurrences. However, it is preferable to use block equi-mean correlation in the Student-t and MGSt copulas rather than one group copula. The block-equivalent models could even be comparable with the multi-group models in all criteria AIC, BIC, and DIC. We also obtain that the group Student-t copula yields lower degrees of freedom than the single group Student-t copula. This finding confirms with Creal and Tsay [2015] due to the fact that when the number of assets in a group increases, the uncertainty reduces because the central limit theorem holds.

In Tables 2.5 and 2.6, we report respectively details of the estimation for the dynamic group

Table 2.4: Estimation results for alternative copula models

	Gaussian block equi (1)	Gaussian 1G (2)	Gaussian multi.group (3)	Student-t block equi (4)	Student-t 1G (5)	Student-t multi.group (6)	MGSt block equi (7)	MGSt 1G (8)	MGSt multi.group (9)
AIC	-163400	-164422	-164658	-192501	-189064	-192197	-197143	-189184	-196901
BIC	-163198	-163627	-163739	-192232	-188262	-191211	-196807	-188377	-195848
DIC	-166377	-167382	-167622	-211507	-208358	-212414	-212257	-208539	-213090
# params	36	142	164	48	143	164	60	144	188
a	[0.032,0.090]	0.070	[0.056,0.146]	[0.023,0.067]	0.026	[0.025,0.067]	[0.025,0.064]	0.026	[0.026,0.069]
b	[0.968,0.997]	0.971	[0.855,0.984]	[0.984,0.999]	0.992	[0.945,0.995]	[0.981,0.999]	0.992	[0.956,0.995]
ν				[6.814,11.907]	11.421	[6.816,11.870]	[7.778,23.391]	11.378	[7.885,23.086]
γ							[-1.230,-0.179]	-0.106	[-1.215,-0.184]
f_c	[1.215,1.858]	[0.957,2.360]	[0.965,2.369]	[1.278,1.997]	[1.286,2.727]	[1.030,2.442]	[1.250,1.986]	[1.288,2.716]	[1.015,2.432]

Posterior estimations for nine one factor copula models and model selection criteria. Three different models are considered (the block-equivalent mean, one group and multiple-group) for three different copula models (Gaussian, Student-t and MGSt). The table reports only the range of the posterior means for the group models and the point estimates for the one group model.

Student-t and dynamic group MGSt copulas. The posterior means of a and b show different dynamic behaviors in each group sector. The values of f_c are depicted as interval range of the posterior means of f_{igc}, for assets $i = 1, \ldots, n_g$ belonging to each group $g = 1, \ldots, G$. The values in parentheses are the average values of the posterior standard deviations. The posterior means of ν are quite different among groups as well as between models. The standard deviations of ν are small in the case of the group Student-t and seem to be higher in the MGSt model. The lowest degree of freedom parameter in the dynamic group Student-t is in Oil industries standing at 6.8. Despite that, it is strongly negative skewed in the MGSt copula model. The other groups that have low degrees of freedom such as Paper, Insurance, and Finance reveal a slight skewness. Although the posterior variance of the degrees of freedom in some industries are higher in the case of group MGSt copula, it still supports for the hypothesis that the lower tail is heavier and the distribution is highly asymmetric rather than there is symmetry in both upper tail and lower tail. We show different tail dependence in each group sector by calculating the average of penultimate tail dependence (Manner and Segers [2011]) of bivariate copulas of the assets belonging to the same sector at quantile 0.5%. The strongest lower tail dependence is 0.280 from the Finance sector with also strong upper tail dependence.

Figure 2.3 describes the posterior mean of the dynamic conditional Kendall-τ correlation among group sectors using the Student and MGSt copula. This posterior mean Kendall-τ between sector g_1 and g_2 is calculated based on the average of the scale parameters over iterations and group

Table 2.5: Results for the group Student-t copula with time-varying factor loadings.

	Oil	Food & Bev.	Pharma.	Plastics	Paper	Steel
a	0.067	0.043	0.059	0.052	0.067	0.034
	(0.005)	(0.007)	(0.011)	(0.014)	(0.006)	(0.004)
b	0.985	0.987	0.977	0.945	0.976	0.986
	(0.002)	(0.005)	(0.009)	(0.037)	(0.004)	(0.003)
ν	6.816	9.827	9.690	8.954	9.292	11.870
	(0.215)	(0.606)	(0.644)	(0.471)	(0.377)	(0.421)
f_c	[1.46,1.75]	[1.17,2.07]	[1.16,1.50]	[1.20,2.12]	[1.19,1.97]	[1.23,2.22]
	(0.098)	(0.085)	(0.076)	(0.056)	(0.076)	(0.070)
# firms	13	7	7	7	12	17
$\lambda_L = \lambda_U$	0.177	0.115	0.098	0.160	0.148	0.151
	Home App.	Electronics	Transportation	Retail	Insurance	Finance
a	0.038	0.025	0.049	0.045	0.041	0.041
	(0.006)	(0.004)	(0.010)	(0.006)	(0.004)	(0.004)
b	0.985	0.995	0.975	0.985	0.988	0.992
	(0.005)	(0.002)	(0.013)	(0.004)	(0.002)	(0.002)
ν	9.457	11.001	9.706	10.659	8.001	7.032
	(0.331)	(0.657)	(0.533)	(0.530)	(0.224)	(0.186)
f_c	[1.33,2.32]	[1.03,1.77]	[1.37,2.03]	[1.20,1.72]	[1.03,2.44]	[1.32,2.44]
	(0.073)	(0.105)	(0.068)	(0.082)	(0.086)	(0.103)
# firms	15	8	7	11	18	18
$\lambda_L = \lambda_U$	0.179	0.122	0.181	0.111	0.194	0.238

Posterior estimations for the interest parameters of the group Student-t factor copula. This includes the posterior means and standard deviations for (a, b, ν) and the values of f_c are depicted as interval range of the posterior means together with the average posterior standard deviations. The tail dependence are calculated as the average of the penultimate tail dependence (Manner and Segers [2011]) of bivariate Student-t copula with the mean correlation of the assets belonging to the same sector at quantile 0.5% (see Appendix A.3).

members as $\frac{1}{n_{g_1} n_{g_2}} \sum_{ij} \rho_{it}\rho_{jt}$ where i, j belong to group g_1 and g_2 respectively, and $i \neq j, g_1 \neq g_2$. And then, we calculate the Kendall-τ correlation equivalent of the factor copula models using Monte Carlo simulation. As we can see, a common pattern is that the correlation increased over time during the subprime mortgage crisis (2008-2009) and the European sovereign debt crisis (peak in 2012). Finance and Insurance sectors suffered most as the correlations go up during the crisis while in other sectors such as Food and Retail, the correlation is less volatile.

Figure 2.4 shows the posterior distribution of the conditional variance and conditional Kendall-t correlation of several companies including Citigroup, Goldman Sachs GP., McDonalds, Johnson & Johnson, Apple, and Intel using MGSt copula. The first two columns illustrate the conditional variance and the last column depicts the conditional Kendall-t correlation between the couple. As mentioned above, the $2007 - 2009$ and $2010 - 2012$ period experienced a high volatility and a

Table 2.6: Results for the group MGSt copula with time-varying factor loadings.

	Oil	Food & Bev.	Pharma.	Plastics	Paper	Steel
a	0.061	0.045	0.066	0.051	0.069	0.035
	(0.006)	(0.008)	(0.014)	(0.013)	(0.006)	(0.004)
b	0.984	0.984	0.970	0.956	0.977	0.986
	(0.003)	(0.006)	(0.013)	(0.023)	(0.004)	(0.003)
ν	23.086	13.967	17.039	10.022	9.875	12.354
	(1.926)	(1.671)	(3.107)	(0.648)	(0.436)	(0.460)
γ	-1.215	-0.450	-0.682	-0.251	-0.236	-0.264
	(0.078)	(0.072)	(0.124)	(0.032)	(0.023)	(0.018)
f_c	[1.38,1.75]	[1.16,2.01]	[1.13,1.48]	[1.18,2.09]	[1.16,1.94]	[1.20,2.18]
	(0.090)	(0.082)	(0.071)	(0.058)	(0.078)	(0.073)
# firms	13	7	7	7	12	17
λ_L	0.222	0.159	0.149	0.204	0.186	0.191
λ_U	0.071	0.065	0.044	0.113	0.103	0.117
	Home App.	Electronics	Transportation	Retail	Insurance	Finance
a	0.034	0.026	0.052	0.047	0.047	0.042
	(0.005)	(0.004)	(0.009)	(0.007)	(0.005)	(0.004)
b	0.989	0.995	0.976	0.985	0.985	0.992
	(0.003)	(0.002)	(0.009)	(0.005)	(0.003)	(0.002)
ν	10.230	12.028	9.997	13.994	8.750	7.885
	(0.399)	(0.819)	(0.629)	(1.051)	(0.296)	(0.297)
γ	-0.257	-0.270	-0.246	-0.418	-0.212	-0.184
	(0.020)	(0.034)	(0.029)	(0.044)	(0.017)	(0.022)
f_c	[1.32,2.29]	[1.01,1.75]	[1.34,1.98]	[1.21,1.71]	[1.02,2.43]	[1.31,2.42]
	(0.082)	(0.110)	(0.070)	(0.082)	(0.081)	(0.104)
# firms	15	8	7	11	18	18
λ_L	0.225	0.157	0.230	0.156	0.239	0.280
λ_U	0.130	0.086	0.139	0.068	0.147	0.184

Posterior estimations for the interest parameters of group MGSt factor copula. This includes the posterior means and standard deviations for (a, b, ν, γ) and the values of f_c are depicted as interval range of the posterior means together with the average posterior standard deviations. The tail dependences are calculated as the average of the penultimate tail dependence (Manner and Segers [2011]) of bivariate MGSt copula with the mean correlation of the assets belonging to the same sector at quantile 0.5%. (see Appendix A.3).

rise in correlation among all examples due to the financial crisis. The cross Kendall-τ correlation between financial series are more volatile than other sectors. They were even highly dependent before crisis happened in 2007.

2.5.3 Risk measures and portfolio allocation

Table 2.7 shows the average of VaR, CVaR and standard deviation of the predicted returns for the equally weighted portfolio. We choose $H = 400$ ahead trading days. All the models perform quite well in terms of risk measure except the one group Student and one group MGSt copulas. In all

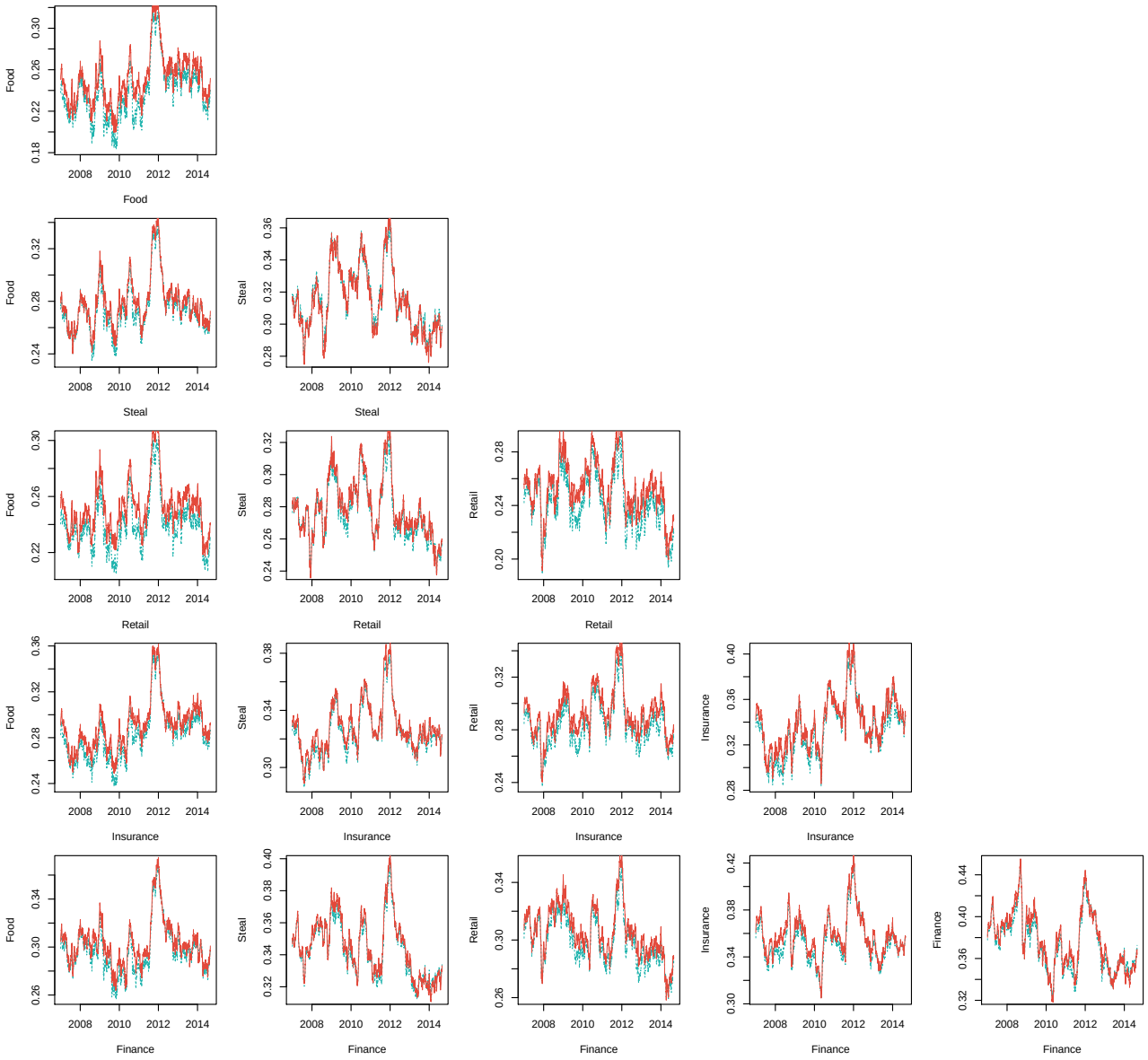

Figure 2.3: The Kendall-τ correlation among group sectors

The figure shows the Kendall-τ correlation among sectors using group Student and MGSt factor copulas. The Kendall-τ correlation for Student model is the blue dash line and Kendall-τ for MGSt model is the red solid line. We select some sectors for illustrating and conserve space. The Kendall-τ correlation among sectors increased during crisis, reached a peak in 2009 and 2012. The online version of this figure is in color.

other cases, the numbers of days that the realized return of the portfolio exceeds the threshold VaR are close to their expected numbers. The block equi-mean correlation and the group MGSt

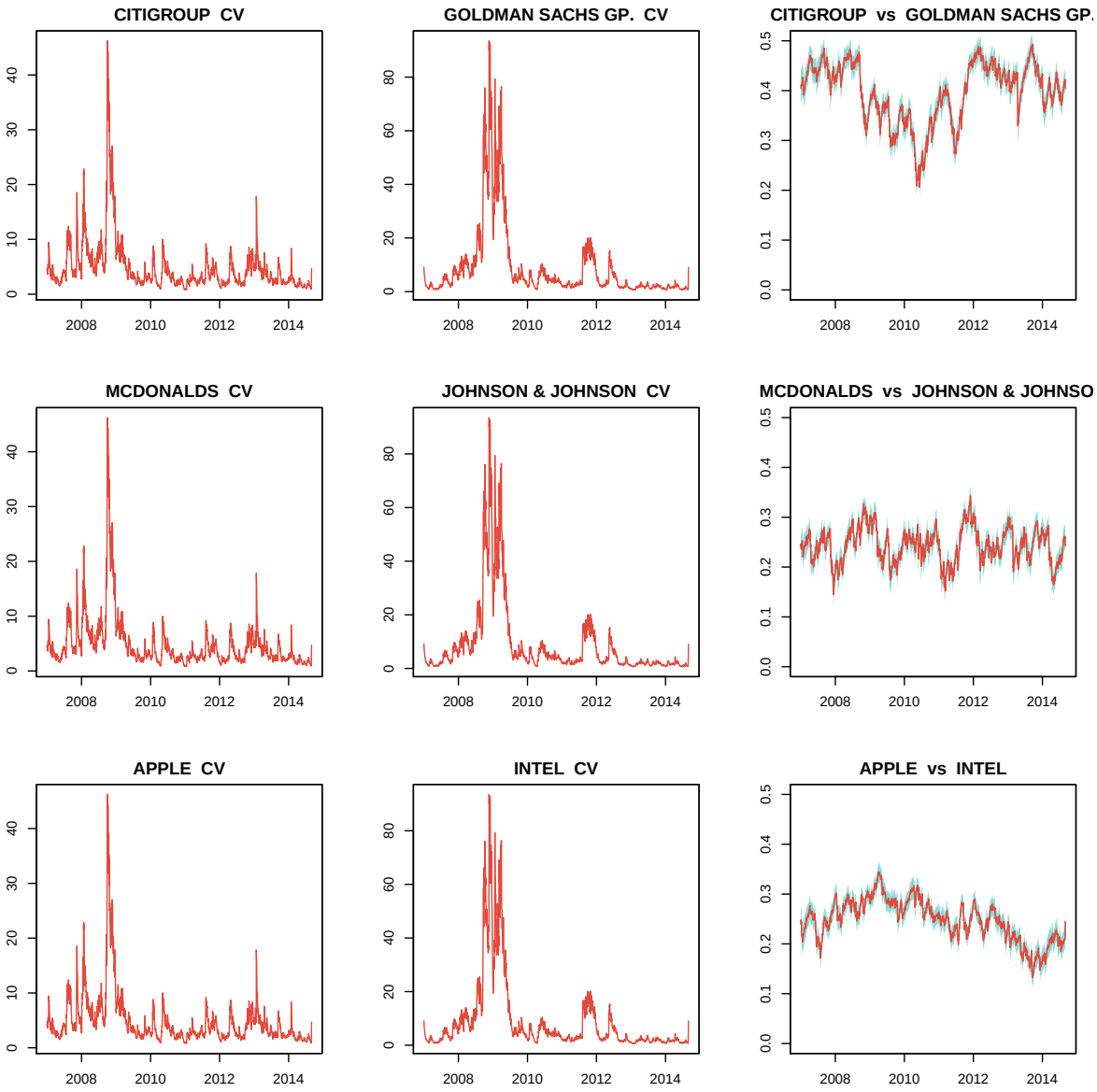

Figure 2.4: Posterior Kendall-τ correlation among time series

The first two columns describe the conditional variance and the last column depicts the dynamic Kendall-τ correlation together with the 95% credible interval using MGSt copula. First row: Citigroup, Goldman Sachs GP., second row: McDonalds, Johnson & Johnson, third row: Apple, Intel. The online version of this figure is in color.

factor copulas also captures correctly quantile dependence at 1% level. The value of CVaR not only depends on the VaR threshold but also depends on the copula types. For copulas that have no or

only one parameter to control for the tail dependence, the CVaR is often higher than those with flexible tail dependence.

Table 2.7: Risk measure for alternative copula models

	Gaussian block equi (1)	Gaussian 1G (2)	Gaussian multi.group (3)	Student-t block equi (4)	Student-t 1G (5)	Student-t multi.group (6)	MGSt block equi (7)	MGSt 1G (8)	MGSt multi.group (9)
VaR_5%	1.664(22)	1.678(18)	1.665(22)	1.617(24)	1.857(16)	1.658(22)	1.59(23)	1.875(19)	1.618(22)
VaR_1%	2.556(4)	2.574(4)	2.556(4)	2.382(5)	3.014(2)	2.456(5)	2.336(5)	3.059(1)	2.379(5)
CVaR_5%	2.237	2.252	2.239	2.108	2.614	2.165	2.059	2.653	2.093
CVaR_1%	3.14	3.152	3.142	2.83	3.879	2.92	2.758	3.96	2.802
Std	1.066	1.074	1.067	1.028	1.212	1.049	1.03	1.21	1.045

Risk measure for nine one factor copula models. The number in the bracket is the number of days that realized return of the portfolio exceeds the threshold VaR. For $H = 400$ trading days, the expected number of violations are 20 days, 4 days at 5%, 1% level correspondingly.

Figure 2.5 shows the smoothed weight for the global optimal portfolio. The figure on the left is obtained by finding the minimum variance portfolio for all assets. We only show the top five assets that have an average weight larger than 5%. The figure on the right illustrates the weight of the minimum CVaR portfolio at 5%. The collections of heavy weight assets chosen in both portfolios are quite similar. The portfolio contains assets that are robust to high volatility such as those in the group Retail, Paper, Pharmaceuticals. We see the similar pattern in both optimal portfolios with holding less "Kellogg" and increasing "Johnson & Johnson", despite that the weights are different. On average, about 60 assets are included in the minimum CVaR in each period while the minimum variance portfolio contains about 15 assets.

2.6 Conclusion

In this chapter, we have proposed a family of one factor copula models and developed a Bayesian algorithm to make parallel inference on the model parameters. In our proposed models, the time series become independent conditioning on the latent factor which allows us to introduce an estimation strategy in a parallel setting. Furthermore, the factor loadings have been modeled as GAS processes which imposes a dynamic dependence structure in their densities. Using multiple-group MGSt copulas, we obtain different types of tail and asymmetric dependence.

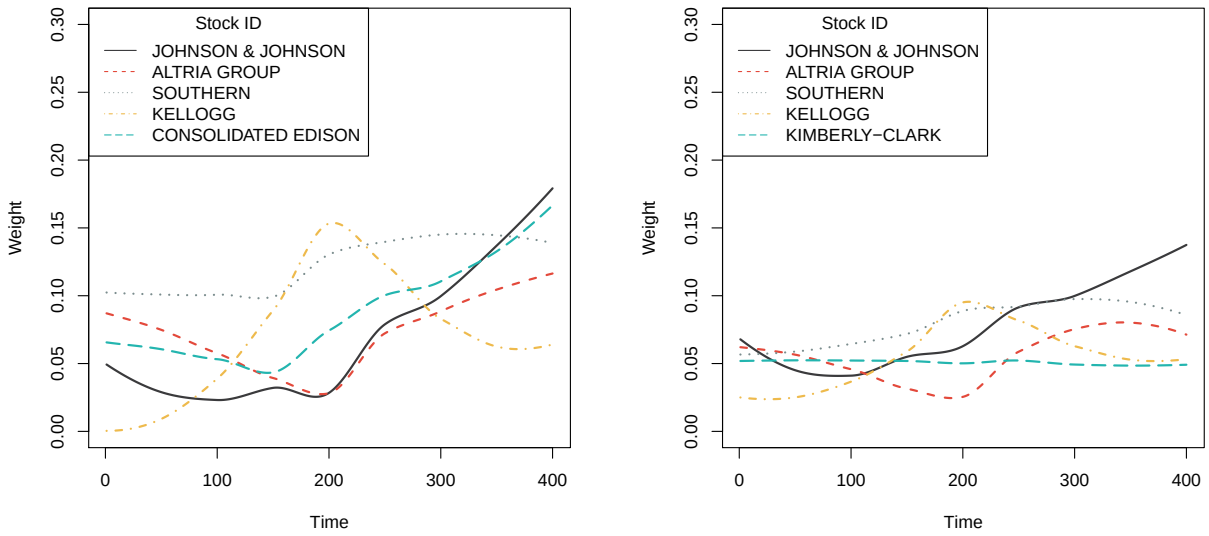

Figure 2.5: Portfolio allocation among time series based on min-variance and min-CVaR

The figure on the left shows the smooth weight for "Johnson & Johnson", "Altria Group", "Southern", "Kellogg", and "Consolidated Edison" in the global minimum variance portfolio. The figure on the right shows the smooth weights on global minimum CVaR portfolio. In the optimal variance and CVaR, there are similar assets that follow a similar trend, the weights are quite different.

The models are extendible since the number of parameters scales linearly with the dimension. As an extension, more complex copula functions can be built based on the distribution of ζ_g. However, this also may require the computational cost to obtain the inverse cdf. Also, we might consider factor models using the family of Archimedean copulas that only allow for lower tail dependence, due to the empirical finding that half of the groups only show weak evidence of upper tail dependence. Finally, one factor models may not be enough for the high dimensional dependence as Oh and Patton [2017a] and Nguyen et al. [2018] suggest. One future direction could be to extend the proposed approach to dynamic multi factor models.

Chapter 3

Structured factor copula models

In this chapter, we propose a Bayesian procedure to make inferences for multi-factor and structured factor copulas in Krupskii and Joe [2013, 2015a]. To deal with the high dimensional structure, we employ a VI algorithm to estimate different specifications of factor copula models. VI aims to approximate the joint posterior distribution of model parameters by a simpler distribution, usually an exponential family. We apply the black-box VI based on the reparameterization of parameters proposed by Kucukelbir et al. [2017]. The central idea is to use several transformations of the factor copula parameters from the constrained space into the real coordinate space and approximate the transformed posterior distribution by a product of univariate Gaussian distributions. Then, VI uses the noisy gradient computed from Monte Carlo simulations to optimize the variational objective. Compared to the MCMC approach, the variational approximation is much faster and could handle a sizeable problem in a few seconds. Furthermore, the posterior means of VI estimation are similar to that of MCMC samples while the standard deviations are only underestimated in the bi-factor copula model. Another issue of factor copula models is that the bivariate copula functions connecting the variables are unknown in high dimensions. We derive an automatic procedure to recover the hidden dependence structure. By taking advantage of the posterior modes of the latent variables, we select the bivariate copula functions based on minimizing the Bayesian information criterion (BIC). All the parameters are estimated jointly hence we take into account the uncertainty

of latent variables as well as copula parameters in each tree layer. Simulation in different context show that the procedures of bivariate copula selection could be at least 80% accurate.

We illustrate the proposed methodology with a high dimensional real data set. We shows an example that the structured factor copula models help to predict the missing temperatures of 24 locations among 479 stations in Germany. In general, the structured factor copula model can capture quite well the dependence structure of high dimensional data. The first common factor reveals most of the tail dependence among the variables with a large proportion of bivariate Student-t copulas. The latent factors in higher layers would correct for the remain dependence with a combination of Gaussian, Frank, and Gumbel copulas.

The rest of the chapter is organized as follows. Section 3.1 introduces the structured factor copula models. Section 3.2 presents the VI algorithm for high dimensional structured factor copula models. Section 3.3 shows some numerical simulations of the proposed factor copula models. Section 3.4 illustrates the applications with real data. Finally, conclusions are reached in Section 3.5.

3.1 Model specification

Let $X = (X_1, \ldots, X_d)'$ be the d-dimensional continuous random variable that we want to model their joint dependence structure and let $F_1(x_1), \ldots, F_d(x_d)$ be their marginal cumulative distribution functions (cdf). Sklar [1959] considers a copula as a joint cdf function defined in the unit hypercube $[0,1]^d$ with uniformly univariate margins, $C(u_1, \ldots, u_d) = F(x_1, \ldots, x_d)$, where $u_i = F_i(x_i)$, for $i = 1, \ldots, d$. It is well-known that the variable $U_i = F_i(X_i) \sim \mathbb{U}(0,1)$, for $i = 1, \ldots, d$. Hence, copula helps to separate the marginal distributions from the dependence structure. Several examples of bivariate copula functions are shown in Table 3.1. However, constructing high dimensional copula functions is difficult. Instead, Bedford and Cooke [2001, 2002], and Aas et al. [2009], among others, decompose the copula density into a sequence of bivariate copulas and conditional bivariate copulas. Therefore, the dependence structure could be considered as a hierarchical vine where the

dependence among variables is driven by bivariate linkages. As the number of variables increases, the number of possible trees as well as the number of vine copula parameters becomes explosive which results in a truncated vine model, see Brechmann et al. [2012].

On the other hand, Krupskii and Joe [2013] and Joe [2014], for tackling the curse of dimensionality, consider several latent variables at the root and describe the dependence structure through bivariate links between copula data and the latent variables, see Figure 3.1. In the next section, we focus on the one factor and structured factor copulas proposed by Krupskii and Joe [2013, 2015a] to model the dependence of variables in high dimensions.

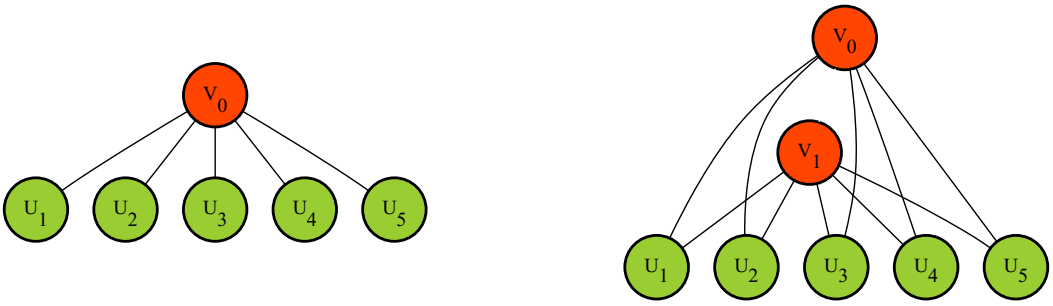

Figure 3.1: One-factor and two-factor copula models (Krupskii and Joe [2013])

Table 3.1: Bivariate copula families and their characteristics

Copula	Notation	Copula distribution function	Prior	Range	Kendall's τ
Gaussian	Gp	$C_{Gp}(u,v;\theta) = \Phi_2(\Phi^{-1}(u), \Phi^{-1}(v); \theta)$	$\pi_{Gp}(\theta) = \pi_{Gn}(\theta) = \frac{2}{\pi}\frac{1}{\sqrt{1-\theta^2}}$	$\theta \in (0,1)$	$\frac{2}{\pi}\texttt{arcsin}(\theta)$
	Gn	$C_{Gn}(u,v;\theta) = \Phi_2(\Phi^{-1}(u), \Phi^{-1}(v); \theta)$		$\theta \in (-1,0)$	
Student-t	Tp	$C_{Tp}(u,v;\theta,\nu) = T_2(T_\nu^{-1}(u), T_\nu^{-1}(v); \theta, \nu)$	$\pi_{Tp}(\theta) = \pi_{Tn}(\theta) = \frac{2}{\pi}\frac{1}{\sqrt{1-\theta^2}}$	$\theta \in (0,1), \nu \in (2,30)$	$\frac{2}{\pi}\texttt{arcsin}(\theta)$
	Tn	$C_{Tn}(u,v;\theta,\nu) = T_2(T_\nu^{-1}(u), T_\nu^{-1}(v); \theta, \nu)$	$\pi_T(\nu) = Gamma(\nu; 1, 0.1)$	$\theta \in (-1,0), \nu \in (2,30)$	
Clayton	C	$C_C(u,v;\theta) = \left(u^{-\theta} + v^{-\theta} - 1\right)^{-\frac{1}{\theta}}$	$\pi_C(\theta) = \pi_{C180}(\theta) = \frac{2}{(\theta+2)^2}$	$\theta \in (0,\infty)$	$\frac{\theta}{\theta+2}$
	C_{180}	$C_{C180}(u,v;\theta) = 1 - u - v + C_C(1-u, 1-v; \theta)$			
	C_{90}	$C_{C90}(u,v;\theta) = v - C_C(1-u, v; -\theta)$	$\pi_{C90}(\theta) = \pi_{C270}(\theta) = \frac{2}{(\theta-2)^2}$	$\theta \in (-\infty,0)$	$\frac{\theta}{\theta-2}$
	C_{270}	$C_{C270}(u,v;\theta) = u - C_C(u, 1-v; -\theta)$			
Gumbel	G	$C_G(u,v;\theta) = \texttt{exp}\left[-\left\{(-\texttt{log}u)^\theta + (-\texttt{log}v)^\theta\right\}^{1/\theta}\right]$	$\pi_G(\theta) = \frac{1}{\theta^2}$	$\theta \in [1,\infty)$	$1 - \frac{1}{\theta}$
	G_{180}	$C_{G180}(u,v;\theta) = 1 - u - v + C_G(1-u, 1-v; \theta)$			
	G_{90}	$C_{G90}(u,v;\theta) = u - C_G(1-u, v; -\theta)$		$\theta \in (-\infty,-1]$	$-1 - \frac{1}{\theta}$
	G_{270}	$C_{G270}(u,v;\theta) = v - C_G(u, 1-v; -\theta)$			
Frank	F_p	$C_{Fp}(u,v;\theta) = -\frac{1}{\theta}\texttt{log}\left(1 - \frac{(1-e^{-\theta u})(1-e^{-\theta v})}{1-e^{-\theta}}\right)$	$\pi_F(\theta) = \frac{4}{\theta^2}(1 - B(\theta) + 2D_1(\theta))$	$\theta \in (0,\infty)$	$1 - \frac{4}{\theta}(1 - D_1(\theta))$
	F_n	$C_{Fn}(u,v;\theta) = -\frac{1}{\theta}\texttt{log}\left(1 - \frac{(1-e^{-\theta u})(1-e^{-\theta v})}{1-e^{-\theta}}\right)$	$\approx Cauchy(\theta; 0, 6)$	$\theta \in (-\infty,0)$	
Joe	J	$C_J(u,v;\theta) = 1 - \left\{(1-u)^\theta + (1-v)^\theta - (1-u)^\theta(1-v)^\theta\right\}^{1/\theta}$	$\pi_J(\theta) = \sum_{k=1}^{\infty} \frac{8(\theta(k-1)+2-1/k)}{(\theta k+2)^2(\theta(k-1)+2)^2}$	$\theta \in [1,\infty)$	$1 - 4\sum_{k=1}^{\infty} \frac{1}{k(\theta k+2)(\theta(k-1)+2)}$
	J_{180}	$C_{J180}(u,v;\theta) = 1 - u - v + C_J(1-u, 1-v; \theta)$	$\approx \frac{2}{(\theta+2)^2}$		
	J_{90}	$C_{J90}(u,v;\theta) = v - C_J(1-u, v; -\theta)$	$\pi_J(\theta) = \sum_{k=1}^{\infty} \frac{8(-\theta(k-1)+2-1/k)}{(\theta k-2)^2(\theta(k-1)-2)^2}$	$\theta \in (-\infty,-1]$	$1 - 4\sum_{k=1}^{\infty} \frac{1}{k(\theta k-2)(\theta(k-1)-2)}$
	J_{270}	$C_{J270}(u,v;\theta) = u - C_J(u, 1-v; -\theta)$	$\approx \frac{2}{(\theta-2)^2}$		
Independence	I	$C_I(u,v) = uv$	-	-	0

$D_1(\theta) = \frac{1}{\theta}\int_0^\theta \frac{\theta}{\texttt{exp}(\theta)-1}$ denotes the Debye function of order one. $B(\theta) = \frac{\theta}{\texttt{exp}(\theta)-1}$ denotes the Bernoulli function.
The table shows some common bivariate copula functions as well as their characteristics such as parameter ranges, and Kendall's τ correlation. We divide the symmetric copula functions into positive and negative Kendall's τ correlation copulas to prevent the identification issue of the factor copula models.

3.1.1 One-factor copula models

In the one-factor copula model proposed by Krupskii and Joe [2013], the dependence structure is characterized through d bivariate copulas between the variable U_i and a latent variable V_0, where $V_0 \sim \mathbb{U}(0,1)$ and $i = 1, \ldots, d$. Let us denote each bivariate copula density by $c_{U_i,V_0}(u_i, v_0; \theta_{0i})$, each bivariate copula function by $C_{U_i,V_0}(u_i, v_0; \theta_{0i})$, and each conditional distribution function of U_i given V_0 by $F_{U_i|V_0}(u_i|v_0; \theta_{0i})$, where θ_{0i} denotes the bivariate copula parameters. Then assuming that the variables $U_1, \ldots, U_d$ are conditionally independent given the latent variable V_0, as shown in Krupskii and Joe [2013], we have

$$F_{U_i|V_0}(u_i|v_0; \theta_{0i}) = \frac{\partial F_{U_i,V_0}(u_i, v_0; \theta_{0i})}{\partial v_0} = \frac{\partial C_{U_i,V_0}(u_i, v_0; \theta_{0i})}{\partial v_0}.$$

where F_{U_i,V_0} is the joint cdf of U_i and V_0. The conditional copula density is the following,

$$\begin{aligned} p(u_1, \ldots, u_d|v_0; \theta) &= \frac{\partial F(u_1, \ldots, u_d|v_0; \theta)}{\partial u_1 \ldots \partial u_d} = \prod_{i=1}^{d} \frac{\partial F_{U_i|V_0}(u_i|v_0; \theta)}{\partial u_i} = \prod_{i=1}^{d} \frac{\partial C_{U_i,V_0}(u_i, v_0; \theta)}{\partial u_i \partial v_0} \\ &= \prod_{i=1}^{d} c_{U_i,V_0}(u_i, v_0|\theta_{0i}), \end{aligned} \tag{3.1}$$

where $\theta = \{\theta_{01}, \ldots, \theta_{0d}\}$ is the set of copula parameters. Note that we have the conditional density $p(u_1, \ldots, u_d|v_0; \theta) = c(u_1, \ldots, u_d|v_0; \theta)$ due to the uniform marginal of U_i, for $i = 1, \ldots, d$. Krupskii and Joe [2013] calculate the unconditional copula density by integrating over the latent space of V_0 and use maximum likelihood to estimate the parameter θ through

$$p(u_1, \ldots, u_d; \theta) = \int_0^1 \prod_{i=1}^{d} c_{U_i,V_0}(u_i, v_0|\theta_{0i}) dv_0.$$

In order to account for more latent variables, Krupskii and Joe [2015a] extend the one-factor copula to the structured factor copulas by adding a hierarchical dependence structure for latent variables, in the case of nested factor copulas or by using the latent variables to capture the conditional

dependence in higher tree layers in the case of bi-factor copulas.

3.1.2 Nested factor copula models

Krupskii and Joe [2015a] propose a nested factor copula model by dividing d variables into G groups, where d_g is the number of variables in group g, for $g = 1, \ldots, G$, such that $\sum_{g=1}^{G} d_g = d$. In each group, the dependence structure is characterized through d_g bivariate copulas between the variable U_{i_g} and the group latent variable V_g, where $V_g \sim \mathbb{U}(0,1)$ and $i_g = 1, \ldots, d_g$. The dependence among groups is determined through G bivariate copulas between the group latent variable V_g and the common latent variable V_0, where $V_0 \sim \mathbb{U}(0,1)$, see Figure 3.2. So, the joint dependence is modelled through $G+1$ latent variables $V = \{V_0, V_1, \ldots, V_G\}'$ and $d+G$ bivariate links. Two variables U_{i_g} and U_{j_g} in the same group are conditionally independent given the latent group factor V_g and they are also conditionally independent from the other group factor $V_{g'}$ for $g \neq g'$, $i_g \neq j_g$. Krupskii and Joe [2015a] consider the nested factor copula model as a hierarchical dependence from a common root variable, hence also as an extension of the one-factor copula model. Let $c_{U_{i_g},V_g}(u_{i_g}, v_g|\theta_{gi_g})$ be the bivariate copula density of U_{ig} and V_g and let $c_{V_g,V_0}(v_g, v_0|\theta_{0g})$ be the bivariate copula density of V_g and V_0, for $g = 1, \ldots, G$, and $i_g = 1, \ldots, d_g$. Then, the conditional density function for the nested factor copulas is the following,

$$p(u_1, \ldots, u_d, v_1, \ldots, v_g|v_0; \theta) = p(u_1, \ldots, u_d|v_1, \ldots, v_g; \theta)p(v_1, \ldots, v_g|v_0; \theta), \tag{3.2}$$

where $\theta = \{\theta_{01}, \ldots, \theta_{0G}, \theta_{11}, \ldots, \theta_{Gd_g}\}$ is the set of copula parameters, and from Eq. (3.1), the conditional density for each layer is,

$$p(u_1, \ldots, u_d|v_1, \ldots, v_g; \theta) = \prod_{g=1}^{G} \prod_{i_g=1}^{d_g} c_{U_{i_g},V_g}(u_{i_g}, v_g|\theta_{gi_g}),$$

$$\text{and } p(v_1, \ldots, v_g|v_0; \theta) = \prod_{g=1}^{G} c_{V_g,V_0}(v_g, v_0|\theta_{0g}).$$

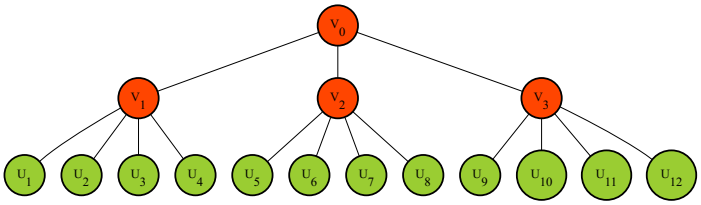

Figure 3.2: Nested factor copulas with $d = 12$ and $G = 3$ (Krupskii and Joe [2015a])

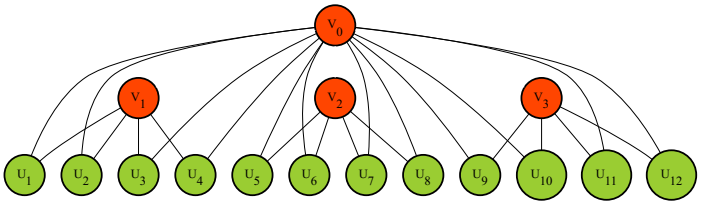

Figure 3.3: Bi-factor copulas with $d = 12$ and $G = 3$ (Krupskii and Joe [2015a])

3.1.3 Bi-factor copula model

Similar to the nested factor copula model, the bi-factor copula model also splits d variables into G groups. However, the dependence of variables is modeled by a sequence of bivariate links between the variables and the latent factors, see Figure 3.3. In fact, bi-factor models extend one-factor copulas by first assuming a one-factor copula in the first tree layer. Then, in the second tree, the relationship among variables in group g is expressed through conditional bivariate copulas between the group latent variable V_g and its member variable U_{i_g} conditional on the common factor V_0, for $i_g = 1, \ldots, d_g$. Krupskii and Joe [2015a] assume that $V_0, \ldots, V_G$ are independent and identically $\mathbb{U}(0,1)$ distributed for identifiability. Therefore, variables in the same group g are conditionally independent given V_0 and V_g, while variables in different groups are conditionally independent given V_0 only. The two-factor copula model in Krupskii and Joe [2013] is a special case of bi-factor copula model where $G = 1$. In order to construct the conditional bi-factor copula density, we follow the pair-copula decomposition method in Aas et al. [2009]. Let $c_{U_{i_g},V_0}(u_{i_g}, v_0|\theta_{0i_g})$ be the bivariate copula density of U_{i_g} and V_0, and let $F_{U_{i_g}|V_0}(u_{i_g}|v_0, \theta_{0i_g})$ be the conditional distribution of U_{i_g} given V_0, for $g = 1, \ldots, G$, and $i_g = 1, \ldots, d_g$. It is straightforward that the conditional variable $U_{i_g|v_0} = F(U_{i_g}|V_0, \theta_{0i_g}) \sim \mathbb{U}(0,1)$. Additionally, let $c_{U_{i_g},V_g|V_0}(u_{i_g|v_0}, v_g|\theta_{gi_g})$ be the conditional

bivariate copula density of $U_{i_g|v_0}$ and V_g given V_0. Therefore, the bi-factor copula density function is given by,

$$\begin{aligned} p(u_1, \ldots, u_d|v_0, \ldots, v_G; \theta) &= \prod_{g=1}^{G} \prod_{i_g=1}^{d_g} p(u_{i_g}|v_0, v_g; \theta) \\ &= \prod_{g=1}^{G} \prod_{i_g=1}^{d_g} c_{U_{i_g}, V_g|V_0}(u_{i_g|v_0}, v_g|\theta_{gi_g}) c_{U_{i_g}, V_0}(u_{i_g}, v_0|\theta_{0i_g}), \end{aligned} \tag{3.3}$$

where $\theta = \{\theta_{01}, \ldots, \theta_{0d}, \theta_{11}, \ldots, \theta_{Gd_G}\}$ is the set of copula parameters.

The bi-factor copula model requires $2d$ bivariate copula links and the computational expensive transformation $u_{i_g|v_0} = F_{U_{i_g}|V_0}(u_{i_g}|v_0, \theta_{0i_g})$ to obtain the transformed variables in the second tree layer. On the other hand, the nested factor copula model only requires $d + G$ bivariate links. Also, there is no need to obtain the conditional variables. According to the simplified assumption for vine copulas in Haff et al. [2010], we also assume that the bivariate copula density $c_{U_{i_g}, V_g|V_0}(u_{i_g|v_0}, v_g)$ does not change with different values of the conditional variable V_0.

3.1.4 Discussion

The structured factor copula models deliver meaningful interpretations on the dependence structure of observable variables. For instance in finance, it is expected that the companies operating in the same country are more dependent and they are all affected not only by a common economic latent factor but also by the country latent factors. The structured factor model could also handle hundreds of variables and capture well different behaviors in the upper and lower tails. In general, the hierarchical structure and group members can be specified based on the prior knowledge as the assumption of hierarchical models. However, the bivariate copula links are unknown. Furthermore, we face several challenges when estimating the factor copula models. Firstly, the choice of bivariate copula links in the one-factor model is arbitrary but we can have two global optimal solutions when estimating a model with symmetric copula families. For example, if we simulate from an one-factor Gaussian copula with all positive correlation

parameters, depending on the initial point, it could happen that all the estimated correlations are negative and the estimated latent variables are equal to the survival function of the true latent. In order to prevent this phenomenon, we divide the copula families into positive dependence copulas and negative dependence copulas. For instance, a positive Gaussian copula has $\theta \in [0, 1]$ and a positive Frank copula has $\theta \in (0, \infty)$ are in the first group, while a negative Gaussian copula $\theta \in [-1, 0)$ and a negative Frank copula $\theta \in (-\infty, 0)$ are in the second group. Secondly, in a higher latent space, without restrictions, Krupskii and Joe [2013] report a circumstance for the two-factor copula model that the dependence of the latent variable in the higher layer is stronger than the first one. They implied a zero correlation for one bivariate link in the second layer to make the model identifiable. Nikoloulopoulos and Joe [2015] reported that the two-factor Student-t copula model is nearly non-identifiable because the likelihood is quite flat. Empirically, when the first tree layer has captured most of the tail dependence, the second tree only reveals a small remaining effect. Hence, we impose not to use the Student-t copula in the higher latent bivariate links and consider other tail dependence families such as the Clayton, Gumbel, and Joe copulas. Thirdly, we also face a situation that the bivariate copula density of Clayton and Survival Joe are very much alike especially when the Kendall's τ correlation is high. This results in a scenario that different factor models containing Clayton and Joe copulas can have similar values of the AIC and/or BIC. Thus, they can be used interchangeable. This phenomenon also happens between the Gaussian and Student copulas where the degree of freedom is high. Finally, Krupskii and Joe [2013] suggest using the bivariate normal score plot or tail-weighted measures of dependence proposed by Krupskii and Joe [2015b] to identify the unknown bivariate links. Because the method requires analyzing each pair of observables, hence becomes infeasible in high dimensions, we propose a procedure in the next section for selecting the best couple links.

3.2 Bayesian approach

In this section, we apply a Bayesian approach to make inferences on both latent variables and bivariate copula parameters. Due to the time demanding cost of MCMC samplers, we employ the VI proposed by Kucukelbir et al. [2017] based on the reparameterization of the model parameters. VI looks for a simple proposal distribution that is closest to the posterior in term of minimizing Kullback-Leibler divergence between them. Firstly, we address how the VI approach makes inferences on the factor copula parameters and the latent variables with known bivariate copula links. Due to several restrictions on the parameters, we transform the constraint space of the copula parameters to the real coordinate space and approximate the transformed posterior distribution in the real domain with a product of Gaussian univariate distributions. Then, we apply the stochastic optimization to minimize the Kullback-Leibler divergence from the transformed posterior distribution to the factorized Gaussian distribution. The unbiased noisy gradient used in optimization is calculated using Monte Carlo samples from the proposal distribution. In the case of unknown copula links, we derive an automatic procedure to select the most appropriate bivariate copula functions. Starting with a random initial structure, we obtain the posterior modes of the proposal distribution of the latent variables. Then, we reassess the agreement of bivariate copulas between the copula data and the latent variables assuming that the latent variables are fixed at the posterior modes. We choose the bivariate copula function that minimizes the BIC for each link in the factor model, that leads to a smaller BIC value of the factor copula model. The procedure is repeated until the bivariate copula functions remain unchanged.

3.2.1 Prior distributions

For all structured factor copula models, we assume $\mathbb{U}(0,1)$ prior distributions for the latent variables $v = (v_0, \ldots, v_G)$. To solve for different restrictions in the bivariate copula parameters, we impose a vague but proper prior distribution for the Kendall's tau correlation such that for positive bivariate copula $\tau \sim \mathbb{U}(0,1)$ and for negative bivariate copula $\tau \sim \mathbb{U}(-1,0)$. Then, we calculate

the induced prior distributions for copula parameters using the transformation of the probability density functions $\pi(\theta) = \pi(\tau)|\frac{\partial \tau}{\partial \theta}|$. Table 3.1 shows the induced prior distributions for the copula parameters θ. For the Student-t copula, we let the degree of freedom $\nu \sim Gamma(1, 0.1)$. In order to simplify the notations, we consider both latent variables and bivariate copula parameters as the parameters of factor models as $\Theta = \{v, \theta\}$. Then, the prior density is $\pi(\Theta) = \pi(v)\pi(\theta)$.

3.2.2 Posterior distributions

We obtain the copula data $u = (u_{t1}, \ldots, u_{td})'$ after fitting the appropriate marginal cdf functions to the original data, $u_{ti} = F_i(x_{ti})$, for $t = 1, \ldots, T$, and $i = 1, \ldots, d$. Assuming that we have specified a factor copula structure together with bivariate links, we are interested in making inferences on the parameters Θ of the structured factor model. The augmented likelihood is,

$$p(u|\Theta) = \prod_{t=1}^{T} p(u_{t1}, \ldots, u_{td}|v_t; \theta).$$

Then, the joint posterior density up to a normalized constant is

$$\pi(\Theta|u) \propto p(u, \Theta) = \prod_{t=1}^{T} p(u_{t1}, \ldots, u_{td}|v_t; \theta) \prod_{t=1}^{T} \pi(v_t)\pi(\theta). \tag{3.4}$$

Substituting Eq. (3.1) into Eq. (3.4) leads to the posterior distribution of Θ for the one-factor copula model,

$$p(\Theta|u) \propto \prod_{i=1}^{d} \left[\prod_{t=1}^{T} c_{U_i,V_0}(u_{ti}, v_{t0}|\theta_{0i})\pi(\theta_{0i}) \right].$$

Similarly, the posterior distribution of Θ for the nested factor copula is given by,

$$p(\Theta|u) \propto \left[\prod_{g=1}^{G} \prod_{i_g=1}^{d_g} \left[\prod_{t=1}^{T} c_{U_{i_g},V_g}(u_{ti_g}, v_{tg}|\theta_{gi_g})\pi(\theta_{gi_g}) \right] \right] \left[\prod_{g=1}^{G} \left[\prod_{t=1}^{T} c_{V_g,V_0}(v_{tg}, v_{t0}|\theta_{0g})\pi(\theta_{0g}) \right] \right],$$

while for the bi-factor copula model is,

$$p(\Theta|u) \propto \prod_{g=1}^{G}\prod_{i=1}^{d_g}\left[\prod_{t=1}^{T} c_{U_{i_g},V_g|V_0}(u_{t,i_g|v_{t0}}, v_{tg}|\theta_{gi_g}) \prod_{t=1}^{T} c_{U_{i_g},V_0}(u_{ti_g}, v_{t0}|\theta_{0i_g})\pi(\theta_{gi_g})\pi(\theta_{0g})\right].$$

The bi-factor copula model requires the conditional transformation $u_{ti_g|v_0} = F(u_{ti_g}|v_{t0})$ whenever we generate new samples of θ_{0i} or v_{t0}, which makes the MCMC approach computationally expensive. However, as we can still calculate the joint posterior distribution of the factor copula parameters, it is feasible to employ the No-U-Turn Sampler (NUTS) proposed by Hoffman and Gelman [2014] to make inferences. In high dimensions, NUTS converges to the target distribution quicker than Metropolis or Gibbs sampling. Even though, we overcome such time demanding sampler by employing an approximated approach to the posterior distribution.

3.2.3 Variational Inference

Considering a factor copula model in Section 3.2.2, the parameter set is Θ, and let N be the total number of parameters in Θ. We assume that the posterior density is approximated by a proposal density $q(\Theta;\lambda)$, parameterized by a vector λ such that $q(\Theta;\lambda)$ is close to the joint posterior $p(\Theta|u)$. In order to find the values of the free parameters in λ, the VI approach tries to minimize the Kullback-Leibler divergence from the posterior $p(\Theta|u)$ to the proposal $q(\Theta;\lambda)$, i.e.:

$$\begin{aligned}\arg\min_{\lambda}\ \mathrm{KL}\left(q(\Theta;\lambda)||p(\Theta|u)\right) &= -\mathbb{E}_q[\texttt{log}p(u|\Theta)] + \mathbb{E}_q[\texttt{log}q(\Theta;\lambda)] \\ \text{such that } supp(q(\Theta;\lambda)) &\subseteq supp(p(\Theta|u))\end{aligned}$$

where the support of the proposal $q(\Theta;\lambda)$ is a subset of the support of the posterior. The proposal $q(\Theta;\lambda)$ needs to be simple for a tractable approximation, and also it should be expressive in order to match closely with the posterior, see Zhang et al. [2018]. Note that, the posterior $p(\Theta|u)$ is only defined up to a normalizing constant, then we can not perform the optimization directly. Instead,

we maximize the evidence lower bound (ELBO) as an equivalent objective function,

$$\begin{aligned}\arg\max_{\lambda} \texttt{ELBO}(q) &= \mathbb{E}_q[\texttt{log}p(u,\Theta)] - \mathbb{E}_q[\texttt{log}q(\Theta;\lambda)] \\ &= \texttt{log}p(u) - KL(q(\Theta;\lambda)||p(\Theta|u)) \leq \texttt{log}p(u)\end{aligned} \tag{3.5}$$

such that when $q(\Theta;\lambda) = p(\Theta|u)$, we have $\texttt{ELBO} = \texttt{log}p(u)$. We obtain $p(u,\Theta) = p(u|\Theta)\pi(\Theta)$ from Eq. (3.4). Following Kucukelbir et al. [2017], we apply an Automatic Differentiation Variational Inference (ADVI) algorithm to maximize the objective function ELBO using noisy estimates of its gradients. Firstly, we transform the constrained parameter space to the real coordinate space using a transformation function $\mathbb{T}_j$ for each parameter, $\tilde{\Theta} = \{\tilde{\Theta}_j\} = \{\mathbb{T}_j(\Theta_j)\} = \mathbb{T}(\Theta)$, for $j = 1, \ldots, N$. Then we assume a product of univariate Gaussian densities as the proposal density,

$$q(\tilde{\Theta};\mu,\sigma^2) = \phi_N(\tilde{\Theta};\mu,\sigma^2) = \prod_{j=1}^{N} \phi(\tilde{\Theta}_j;\mu_j,\sigma_j^2), \tag{3.6}$$

where $\mu = \{\mu_j\}$ and $\sigma = \{\sigma_j\}$ for $j = 1, \ldots, N$. This factorized distribution is also called a mean field distribution. Kucukelbir et al. [2017] propose several transformation functions $\mathbb{T}$ to suit with different restrictions in the domain of Θ, see Table 3.2. Let $\omega = \{\omega_j\} = \texttt{log}(\sigma)$ to relax the positive constraint of σ, the variational parameters become $\lambda = \{\mu_j, \omega_j\}$, for $j = 1, \ldots, N$. The joint density $p(u,\tilde{\Theta})$ is derived based on the Jacobian of the inverse transformation, $J_{\mathbb{T}^{-1}}(\tilde{\Theta})$,

$$p(u,\tilde{\Theta}) = p(u,\mathbb{T}^{-1}(\tilde{\Theta}))|\det J_{\mathbb{T}^{-1}}(\tilde{\Theta})|. \tag{3.7}$$

So, instead of optimizing in the constrained parameter space, we optimize the variational parameters in the real space $\mathcal{R}^N$. Substituting Eq. (3.7) into the ELBO Eq. (3.5), the objective function becomes,

$$\arg\max_{\lambda} \texttt{ELBO}(q) = \arg\max_{\lambda=\{\mu,\omega\}} \mathbb{E}_{q(\tilde{\Theta})}\left[\texttt{log}p(u,\mathbb{T}^{-1}(\tilde{\Theta})) + \texttt{log}|\det J_{\mathbb{T}^{-1}}(\tilde{\Theta})|\right] - \mathbb{E}_{q(\tilde{\Theta})}[\texttt{log}q(\tilde{\Theta};\lambda)] \tag{3.8}$$

where $\mathbb{E}_{q(\tilde{\Theta})}[-\texttt{log}q(\tilde{\Theta};\lambda)] = \sum_{j=1}^{N} 0.5\texttt{log}(2\pi e(\exp\omega_j)^2)$ is the closed-form formula of the entropy of multivariate Gaussian distribution. For each factor copula model, the objective function ELBO corresponds to the joint posterior density up to a normalized constant $p(u,\Theta)$. The Monte Carlo method is used to evaluate the ELBO by sampling S samples $\tilde{\Theta}_{(s)} \sim \Phi_N(\mu,(\exp\omega)^2)$ and plugging Eq. (3.4) into Eq. (3.8) as,

$$\texttt{ELBO}(q) \approx \frac{1}{S}\sum_{s=1}^{S}\left[\texttt{log}p(u,\mathbb{T}^{-1}(\tilde{\Theta}_{(s)})) + \texttt{log}|\det J_{\mathbb{T}^{-1}}(\tilde{\Theta}_{(s)})|\right] - \mathbb{E}_{q(\tilde{\Theta})}[\texttt{log}q(\tilde{\Theta};\lambda)]. \tag{3.9}$$

We use the stochastic gradient ascent to maximize the ELBO over $\lambda = \{\mu,\omega\}$. The gradients of ELBO with respect to parameters λ of the proposal distribution is denoted as ∇_λELBO. Using a similar trick, the noisy gradient is approximated by drawing M samples $\tilde{\Theta}_{(m)} = \mu + \exp(\omega)\eta_{(m)}$, where $\eta_{(m)} \sim \Phi(0, I_N)$, and taking the average over the gradients of each sample point. Kucukelbir et al. [2017] recommend that $M = 1$ is sufficient for this purpose. The derivatives of the ELBO with respect to λ follow the chain rule when we apply the transformations,

$$\begin{aligned}
\nabla_\lambda\texttt{ELBO} =& \nabla_\lambda\mathbb{E}_{q(\tilde{\Theta})}\left[\texttt{log}p(u,\mathbb{T}^{-1}(\tilde{\Theta})) + \texttt{log}|\det J_{T^{-1}}(\tilde{\Theta})|\right] - \nabla_\lambda\mathbb{E}_{q(\tilde{\Theta})}[\texttt{log}q(\tilde{\Theta};\lambda)] \\
\approx& \frac{1}{M}\sum_{m=1}^{M}\nabla_\lambda\left[\texttt{log}p(u,\mathbb{T}^{-1}(\tilde{\Theta}_{(m)})) + \texttt{log}|\det J_{T^{-1}}(\tilde{\Theta}_{(m)})|\right] - \nabla_\lambda\mathbb{E}_{q(\tilde{\Theta})}[\texttt{log}q(\tilde{\Theta};\lambda)] \\
\approx& \frac{1}{M}\sum_{m=1}^{M}\left[\nabla_\Theta\texttt{log}p(u,\Theta_{(m)})\nabla_{\tilde{\Theta}}\mathbb{T}^{-1}(\tilde{\Theta}_{(m)}) + \nabla_{\tilde{\Theta}}\texttt{log}|\det J_{T^{-1}}(\tilde{\Theta}_{(m)})|\right]\nabla_\lambda\tilde{\Theta}_{(m)} - \nabla_\lambda\mathbb{E}_{q(\tilde{\Theta})}[\texttt{log}q(\tilde{\Theta};\lambda)]
\end{aligned} \tag{3.10}$$

Schepsmeier and Stöber [2014] derive the partial derivatives of a copula density $c(u,v|\theta)$ w.r.t. its arguments. Hence, it is straightforward to obtain $\nabla_\Theta\texttt{log}p(u,\Theta_{(m)})$ and ∇_λELBO. Then, the free variational parameter λ is updated along the gradient, $\lambda \leftarrow \lambda + \varrho\nabla_\lambda$ELBO with a vector step size ϱ. The vector of step size sequence or the learning rate ϱ needs to be adaptively adjusted corresponding to the variance of the gradient. When the variance of the gradient is large, we expect the step size to be small and vice-versa. Kucukelbir et al. [2017] modify the RMSPROP sequence in

Tieleman and Hinton [2012] to guarantee that the step size consequence decays sufficiently, see Appendix B.1.

Table 3.2: Transformation functions from a constraint domain to the real domain

Parameter range	$\tilde{\Theta} = \mathbb{T}(\Theta) \in \mathbb{R}$	$\Theta = \mathbb{T}^{-1}(\tilde{\Theta})$	$J_{\mathbb{T}^{-1}}(\tilde{\Theta}) = \frac{\partial \mathbb{T}^{-1}(\tilde{\Theta})}{\partial \tilde{\Theta}}$
$\theta \in [0, 1]$	$\tilde{\theta} = \texttt{log}\left(\frac{\theta}{1-\theta}\right)$	$\theta = \frac{\texttt{exp}\tilde{\theta}}{1+\texttt{exp}\tilde{\theta}}$	$J = \frac{\texttt{exp}\tilde{\theta}}{(1+\texttt{exp}\tilde{\theta})^2}$
$\theta \in [0, \infty]$	$\tilde{\theta} = \texttt{log}\left(\theta\right)$	$\theta = \texttt{exp}\tilde{\theta}$	$J = \texttt{exp}\tilde{\theta}$
$\theta \in [L, \infty]$	$\tilde{\theta} = \texttt{log}\left(\theta - L\right)$	$\theta = \texttt{exp}\tilde{\theta} + L$	$J = \texttt{exp}\tilde{\theta}$
$\theta \in [-1, 0]$	$\tilde{\theta} = \texttt{log}\left(\frac{-\theta}{1+\theta}\right)$	$\theta = -\frac{\texttt{exp}\tilde{\theta}}{1+\texttt{exp}\tilde{\theta}}$	$J = -\frac{\texttt{exp}\tilde{\theta}}{(1+\texttt{exp}\tilde{\theta})^2}$
$\theta \in [-\infty, 0]$	$\tilde{\theta} = \texttt{log}\left(-\theta\right)$	$\theta = -\texttt{exp}\tilde{\theta}$	$J = -\texttt{exp}\tilde{\theta}$
$\theta \in [-\infty, U]$	$\tilde{\theta} = \texttt{log}\left(U - \theta\right)$	$\theta = U - \texttt{exp}\tilde{\theta}$	$J = -\texttt{exp}\tilde{\theta}$
$\theta \in [L, U]$	$\tilde{\theta} = \texttt{log}\left(\frac{\theta-L}{U-\theta}\right)$	$\theta = \frac{L+U\texttt{exp}\tilde{\theta}}{1+\texttt{exp}\tilde{\theta}}$	$J = \frac{(U-L)\texttt{exp}\tilde{\theta}}{(1+\texttt{exp}\tilde{\theta})^2}$
$\theta \in [-\infty, \infty]$	$\tilde{\theta} = \theta$	$\theta = \tilde{\theta}$	$J = 1$

Transformation function for different restrictions in constrain parameter space of Θ

Algorithm 1 outlines the procedure of ELBO optimization. Starting with an initial copula structure and the Gaussian proposal distribution. We sample from the proposal and obtain the values $\tilde{\Theta}_{(m)}$. After that, we calculate the noisy gradient of ELBO w.r.t. to the variational parameter λ in Eq. (3.10), and update them along the gradient line using adaptive step sizes. We calculate the value of ELBO using Eq. (3.9) and stop if the relative change is less than a threshold. We recover the distribution of Θ using the inverse transformation $\Theta = \mathbb{T}^{-1}(\tilde{\Theta})$ when sampling $\tilde{\Theta} \sim \Phi_N(\mu, exp(\omega)^2)$. In general, the variational distribution of Θ is non-Gaussian due to the Jacobian transformation.

Data: Copula data $u = \{u_{ti}\}$ and a structured copula model

Result: Bivariate copula links, and samples of Θ from the proposal distribution

Initial bivariate copula links;

while *Any change in copula types* **do**

  Initialize $i = 0$, vector $\mu^{(i)} = 0, \omega^{(i)} = 0$;

  while *Change in* `ELBO` *is above some threshold* **do**

    Draw M samples $\eta_m \sim \Phi(0, I_N)$;

    Obtain $\tilde{\Theta}_{(m)} = \mu^{(i)} + \exp(\omega^{(i)})\eta_{(m)}$;

    Obtain the noisy gradient ∇_μ`ELBO` and ∇_ω`ELBO` ;

    Update $\mu^{(i+1)} \leftarrow \mu^{(i)} + \varrho_\mu^{(i)} \nabla_\mu$`ELBO` ;

    Update $\omega^{(i+1)} \leftarrow \omega^{(i)} + \varrho_\omega^{(i)} \nabla_\omega$`ELBO` ;

    Incremental iteration (i) ;

  end

  Select bivariate copula links between u and $\bar{v}$ based on minimum the BIC ;

  Reassign the copulas and estimate ;

end

Sample $\tilde{\Theta} \sim \Phi_N(\mu, exp(\omega))$, obtain $\Theta = \mathbb{T}^{-1}(\tilde{\Theta})$;

Return bivariate copula links and Θ samples ;

Algorithm 1: Modified of the ADVI algorithm in [Kucukelbir et al., 2017]

3.2.4 Model check

Given a structured factor copula, it is straightforward to derive the `ELBO` and optimize using the ADVI algorithm. However, when the bivariate copula links are unknown, we can take advantage of the posterior modes of the latent variables, called $\bar{v}$, to inspect the assumption of the initial links. The idea is to find the bivariate copula functions that minimize the BIC of bivariate copulas between u and $\bar{v}$. In particular, we start with a random initial structure and obtain the posterior

modes $\bar{v}$. Then, we reassess the agreement of bivariate copulas between u_i and $\bar{v}$. The ideal bivariate function should minimize the BIC of each couple link. By selecting these functions, the model goodness of fit increases. For example, in one-factor copula model, let $c_{U_i,V_0}(u_i, v_0; \theta_{0i})$ be the current copula density of the link between variable U_i and V_0. The BIC for this bivariate link is,

$$\texttt{BIC}_i = -2\texttt{log}c_{U_i,V_0}(u_i, \bar{v}_0; \hat{\theta}_{0i}) + n_i\texttt{log}(T)$$

where $\bar{v}_0 = \{\bar{v}_{10}, \ldots, \bar{v}_{T0}\}$ is the posterior modes of the latent variables, $\hat{\theta}_i$ is the maximum likelihood estimator of the bivariate copula, and n_i is the number of parameters in c_{U_i,V_0}. The BIC of the one factor copula could be derived as,

$$\begin{aligned}\texttt{BIC} &= -2\sum_{i=1}^{d}\texttt{log}c_{U_i,V_0}(u_i, \bar{v}_0; \hat{\theta}_{0i}) + (T + \sum_{i=1}^{d} n_i)\texttt{log}(T) \\ &= \sum_{i=1}^{d}\texttt{BIC}_i + T\texttt{log}(T).\end{aligned}$$

In the selection step, we propose a new bivariate link $c^{(*)}_{U_i,V_0}(u_i, v_0; \theta_{0i})$ among possible copula functions (see Table 3.1) which minimizes the bivariate copula BIC,

$$\underline{\texttt{BIC}}_i^{(*)} = -2\texttt{log}c^{(*)}_{U_i,V_0}(u_i, \bar{v}_0; \theta^{(*)}_{0i}) + n_i^{(*)}\texttt{log}(T) \leq \texttt{BIC}_i$$

where $n_i^{(*)}$ is the number of parameters and $\theta^{(*)}_{0i}$ is the maximum likelihood estimation of $c^{(*)}_{U_i,V_0}$. Therefore,

$$\underline{\texttt{BIC}}^{(*)} = \sum_{i=1}^{d}\underline{\texttt{BIC}}_i^{(*)} + T\texttt{log}(T) \leq \sum_{i=1}^{d}\texttt{BIC}_i + T\texttt{log}(T) = \texttt{BIC}$$

If there are any changes in the bivariate copula links, we update the copula structure and estimate the new model. The $\texttt{BIC}^{(*)}$ of the new model will be at least as good as the $\underline{\texttt{BIC}}^{(*)}$. This guarantees that the model goodness of fit increases. Until there is no difference in the bivariate copula selection, we reach convergence. Accordingly, the posterior modes of the latent variables help to find out

the most appropriate candidate for each bivariate links. For the bi-factor models, one can perform bivariate copula selection jointly for all tree layers, however it requires exponential computation. Here, we prioritize the bivariate copula selection of the first tree at root. Then we obtain the transformation $u_{i_g|v_0}$ and implement the selection procedure for higher layers. In Section 3.3, we illustrate this trade-off for accuracy in different simulated contexts. Algorithm 1 summarizes the procedure for selecting bivariate copula links in the factor copula models.

3.3 Data simulation

In this section, we illustrate the VI algorithm with simulated data sets. We generate samples of $d = 100$ time series with $T = 1000$ observation points. First, we randomize the latent variable v uniformly in the unit range $[0, 1]$, then we simulate copula data u based on the conditional bivariate copulas of u_i and v, for $i = 1, \ldots, d$. The chosen bivariate links are all Gaussian copulas, all Student-t copulas, all Clayton copulas, all Gumbel copulas, all Frank copulas, all Joe copulas and mixing among bivariate copulas. The experiments for the rotated Archimedean copulas have the similar results. The Kendall's τ correlation of each bivariate copula link is randomized uniformly in the range $[0.2, 0.8]$.

In the first experiment, we estimate the factor models given their known structure, while in the second experiment, we start with a random bivariate copula structure and let Algorithm 1 search for the most appropriate bivariate links in the factor copula models. We consider the copula families shown in Table 3.1 that contains 7 common bivariate copula functions and their rotations. We repeat the experiment 100 times and calculate the average of statistical criteria. We report the time in seconds using one Intel Core i7-4770 @ 3.40GHz processor. The computation is implemented in the `vifcopula` package.

3.3.1 One-factor copula model

We report a comparison between the posterior means using variational approximation and the true generated values, see Figure 3.4. In general, the VI posterior means are close to their true values. The posterior means of the degree of freedom parameters in the Student-t factor copulas are less reliable when $\nu > 10$. This is due to the fact that the likelihood of the Student copula is quite flat when ν is high which makes the posterior samples of ν right skewed. Also in general, the posterior medians or posterior modes are closer to the true values. We also compare with the posterior samples from the MCMC approach and the results are similar. Table 3.3 shows the summary of statistical criteria for the one-factor Gaussian, Student-t, Clayton, Gumbel, Frank, Joe and Mix copulas. The result of the first experiment is shown in Panel (a) and that of the second experiment is shown in Panel (b). Each factor copula model contains 100 bivariate links with about 100 to 200 copula parameters. We use Gauss-Legendre quadrature integration over the latent space to obtain $\texttt{log}p(u|\theta)$. The value of ELBO, AIC, BIC, $\texttt{log}p(u|\theta)$ are normalized for 1000 data observations. It usually took less than a minute to estimate the copula parameters at the correct structure model, more details are reported later in Section 3.3.4. In the second experiment, we start with random bivariate links as an initial dependence structure and let Algorithm 1 recover the original structure, it takes a few iterations of bivariate copula selection to converge. The accuracy rate is quite high, except for the Clayton and Joe. Bivariate copula selection often goes wrong between the Clayton copula and the survival Joe copula when the Kendal's τ is high and vice versa. Due to that reason, we restrict the maximum number of selection iterations under 10. After all, we still archive a decent level of the ELBO, AIC and BIC even with misspecified models. The time of estimation gets longer in the case of unknown structure mainly because we need to perform the bivariate copula selection for $d = 100$ bivariate links across all possible copula functions. In general, given copula data u and the estimated latent factor v_0, we need to calculate the BIC of d bivariate links for 18 copula types and select the best functions. It is about one second for each bivariate selection, and one can parallel the procedure to speed up the calculation.

Table 3.3: Model comparison for the one-factor copula models

Copula type	Gaussian	Student-t	Clayton	Gumbel	Frank	Joe	Mix
(a) Initial at the correct structure							
ELBO	31.3	32.6	75.2	67.9	56.6	77.1	56.2
AIC	-63.2	-65.5	-146.4	-134.8	-114.3	-149.9	-111.5
BIC	-62.7	-64.5	-146.0	-134.3	-113.8	-149.4	-110.9
$\texttt{log}p(u\|\theta)$	31.7	32.9	73.3	67.5	57.2	75.1	55.9
(b) Initial at a random structure							
# Selection iteration	3	5	10	2	3	10	7
% accuracy	99	80	70	99	99	61	85
ELBO	31.3	32.6	75.2	67.9	56.6	77.2	56.2
AIC	-63.2	-65.5	-146.5	-134.8	-114.3	-149.9	-111.5
BIC	-62.7	-64.6	-146.0	-134.3	-113.8	-149.5	-111.0
$\texttt{log}p(u\|\theta)$	31.7	32.9	73.3	67.5	57.2	75.1	55.9

We report the statistical criteria for the one-factor copula models. Each factor copula model contains 100 bivariate links with about 100 to 200 copula parameters. We use Gauss-Legendre quadrature integration over the latent space to obtain $\texttt{log}p(u|\theta)$. The value of ELBO, AIC, BIC, $\texttt{log}p(u|\theta)$ are normalized for 1000 data observations.

3.3.2 Nested factor copula model

We randomly divide $d = 100$ time series into $G = 5$ groups and simulate data from the nested factor copulas. Figure 3.5 compares the posterior means of the nested factor model to the true generated values. The estimates of common factor v_0 fluctuate significantly around the true means which brings a lot of uncertainty. Note that in this case, we have only 5 group latent variables to infer the distribution of their common factor. Hence, the estimates will be improved when increasing the number of latent groups. We also expect the selection of bivariate links among the latent variables to be less accurate than that between the observable variables and the latent group variables. Table 3.4 shows the summary of statistical criteria for nested factor copulas. In the second experiment with a random initial structure, it also took a few iterations of bivariate copula selection to converge. Again, the similarity of Clayton and survival Joe copula reduces the number of correct specified links, as shown in Panel (b). The accuracy of bivariate copula functions between observable variables and group latent factors is still high. If the bivariate copulas are

mixed, we still have the accuracy rate at 83%. The time of inference is still quite comparable with the one-factor model, even in this case we have much more latent parameters. More details are shown in Section 3.3.4.

Table 3.4: Model comparison for the nested factor copula models

Copula type	Gaussian	Student-t	Clayton	Gumbel	Frank	Joe	Mix
(a) Initial at the correct structure							
ELBO	25.9	27.9	69.3	61.3	50.7	70.3	49.7
AIC	-52.9	-56.8	-137.2	-122.9	-103.5	-139.1	-99.9
BIC	-52.3	-55.8	-136.7	-122.4	-103.0	-138.5	-99.3
$\texttt{log}p(u\|\theta)$	26.5	28.6	68.7	61.6	51.8	69.6	50.1
(b) Initial at a random structure							
# Selection iteration	4	5	10	3	4	10	8
% accuracy	96	77	75	99	97	52	83
ELBO	25.8	27.8	69.3	61.2	50.6	70.3	49.7
AIC	-52.7	-56.7	-137.0	-122.8	-103.2	-138.9	-99.8
BIC	-52.2	-55.8	-136.5	-122.3	-102.7	-138.3	-99.2
$\texttt{log}p(u\|\theta)$	26.5	28.5	68.6	61.5	51.7	69.5	50.0

We report the statistical criteria for the nested factor copula models. Each factor copula model contains 6 latent factors, 105 bivariate links with about 105 to 210 copula parameters. We use Gauss-Legendre quadrature integration over the latent space to obtain $\texttt{log}p(u|\theta)$. The value of ELBO, AIC, BIC, $\texttt{log}p(u|\theta)$ are normalized for 1000 data observations.

3.3.3 Bi-factor copula model

Similar to the nested factor copula experiments, we randomly divide $d = 100$ variables into $G = 5$ groups. For the identification issue of bi-factor Student-t copulas, the bivariate copulas in the second tree are mixed of other copula families, see Section 3.1.4. Figure 3.6 in Appendix **??** shows the posterior mean using VI approximation in comparison to the true generated values. The posterior means of v_0 are close to their true values while that of v_g vary in a greater range. Note that, all information of variables is used to make inference on the common latent variable v_0, hence we obtain more certainty of v_0 than the group latent variables v_g. Computational time significantly increases because we need to obtain the intensive transformation of $u_{i_g|v_0} = F(u_{i_g}|v_0)$ when calculating ELBO and its derivatives. As a consequence, if the bivariate links in the first tree

are not correctly specified, the links in the second tree would likely be misspecified. Table 3.5 shows the statistical criteria of bi-factor copula models. We obtain quite a good accuracy in both tree layers. Except for the bi-factor Joe copulas, we could at least recover 70% of bivariate links in bi-factor copula models. The statistical criteria are quite close when the correct structure and the recovered structure are compared. Even in this case, we perform the copula selection separately for each tree layer. We prioritize the copula selection in the first layer and obtain the conditional variable $u_{i_g|v_0}$. Then, we obtain the copula selection for the copula links between $u_{i_g|v_0}$ and v_g in the second layer.

Table 3.5: Model comparison for the bi-factor copula models

Copula type	Gaussian	Student-t	Clayton	Gumbel	Frank	Joe	Mix
(a) Initial at the correct structure							
ELBO	56.2	83.8	140.9	126.2	105.0	143.2	107.6
AIC	-115.1	-170.4	-275.0	-254.4	-214.7	-279.4	-212.7
BIC	-114.1	-168.9	-274.0	-253.4	-213.7	-278.5	-211.6
log$p(u\|\theta)$	57.8	85.5	137.7	127.4	107.5	139.9	106.6
(b) Initial at a random structure							
# Selection iteration	4	9	9	5	9	10	9
% accuracy Tree 1	99	76	84	99	99	46	85
% accuracy Tree 2	97	83	19	66	92	2	67
ELBO	56.2	83.8	132.2	124.0	104.8	134.2	105.5
AIC	-115.1	-170.1	-263.1	-250.6	-214.0	-266.8	-209.7
BIC	-114.1	-168.8	-262.1	-249.6	-213.0	-265.8	-208.7
log$p(u\|\theta)$	57.7	85.3	131.7	125.5	107.2	133.6	105.1

We report the statistical criteria for the bi-factor copula models. Each factor copula model contains 6 latent factors, 200 bivariate links with about 200 to 300 copula parameters. We use Gauss-Legendre quadrature integration over the latent space to obtain log$p(u|\theta)$. The value of ELBO, AIC, BIC, log$p(u|\theta)$ are normalized for 1000 data observations.

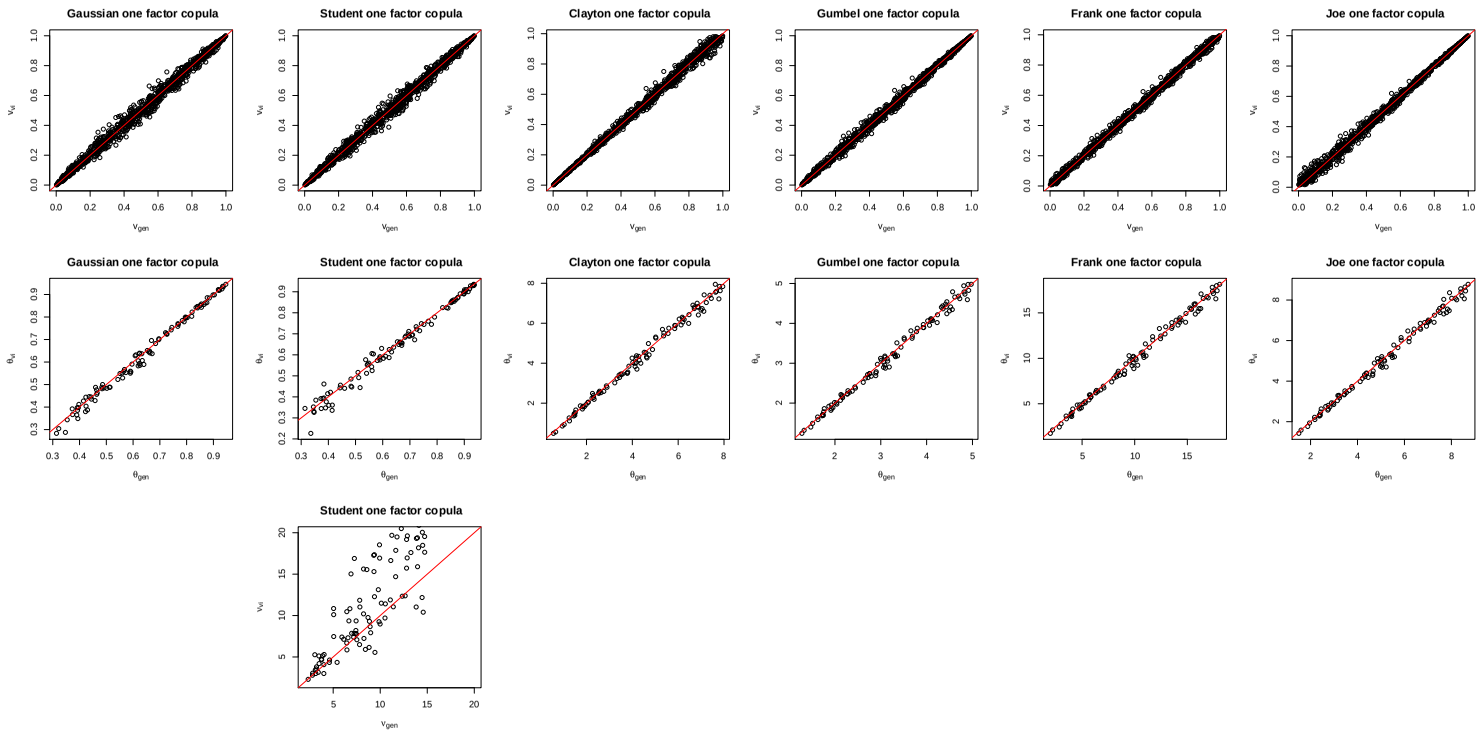

Figure 3.4: Variational inference for the one-factor copula models.

The figure compares the posterior means using variational approximation to the true generated values of the one factor copula models. In general, the posterior means are close to its true generated values. The posterior means of the degree of freedom in the Student-t factor copula are less reliable when $\nu > 10$ due to the flat likelihood.

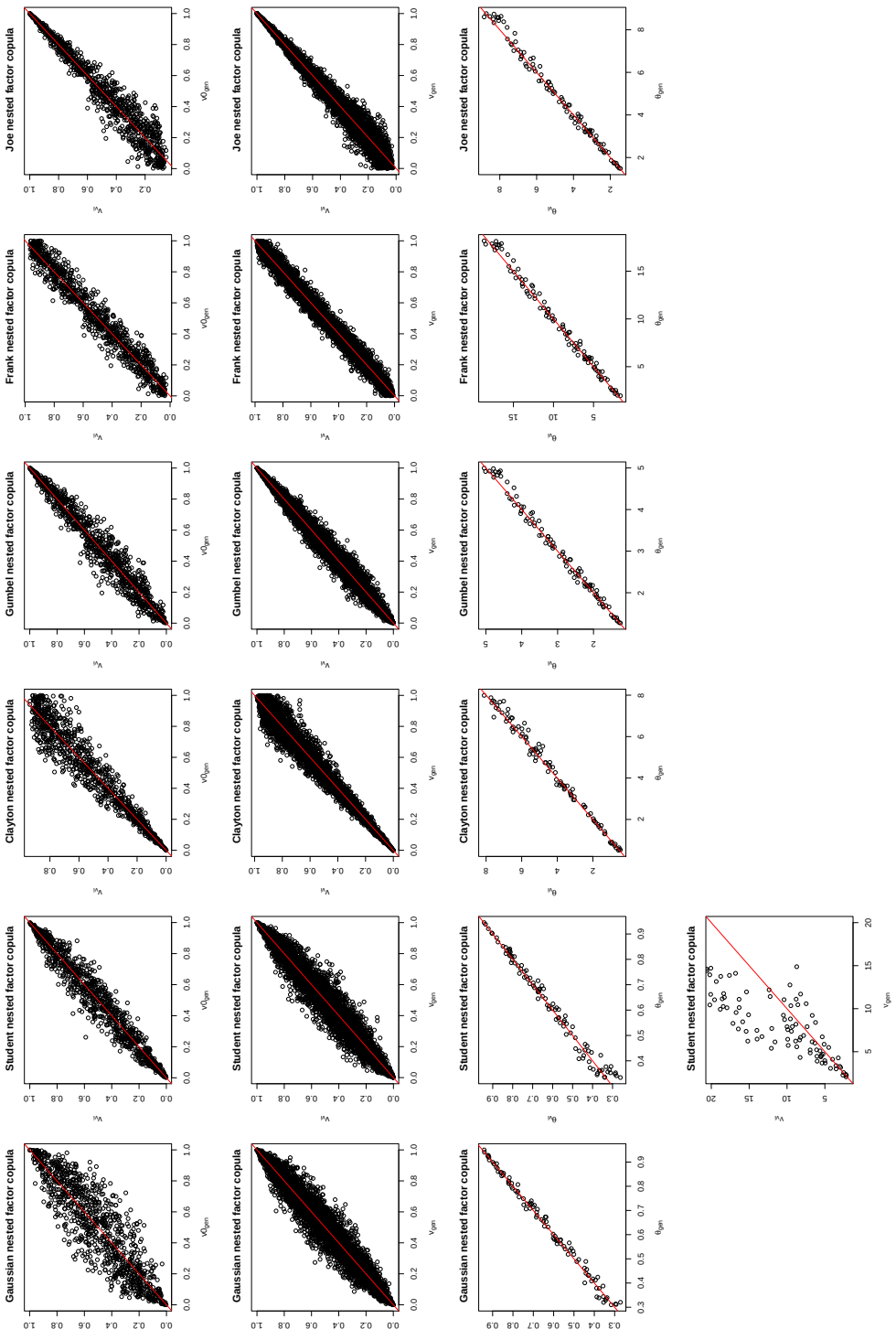

Figure 3.5: Variational inference for the nested factor copula models.

The figure compares the posterior means using variational approximation to the true generated values for the nested factor copula models. The common factor v_0 fluctuates significantly around the true mean due to a small number of group latent variables.

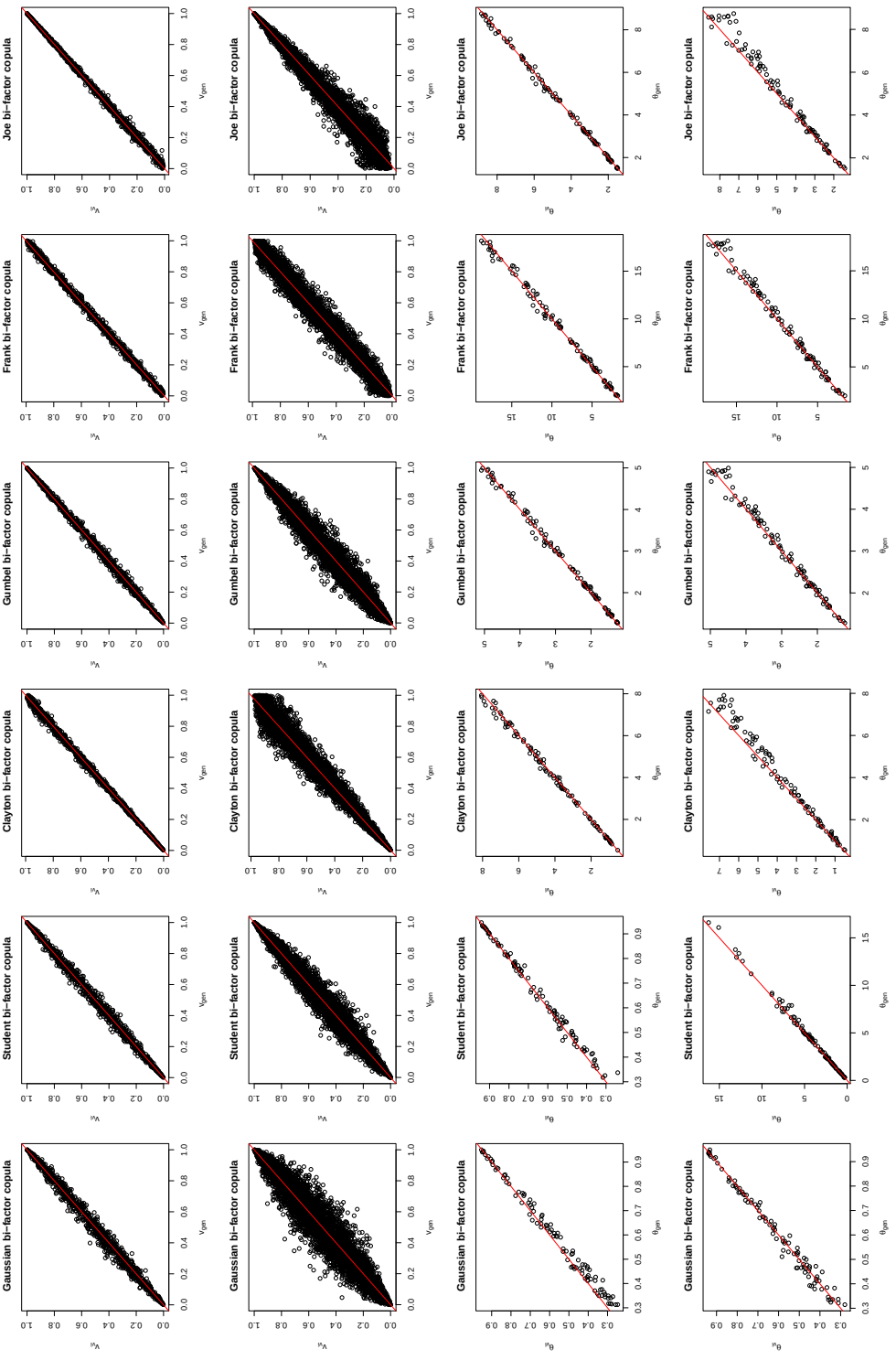

Figure 3.6: Variational inference for the bi-factor copula models.

The figure compares the posterior means using variational approximation to the true generated values of the bi-factor copula models. For the bi-factor Student-t copula, the bivariate copulas in the second tree are mixed of other copula families due to the identification issue.

3.3.4 Comparison between VI and MCMC estimation

We incorporate the NUTS algorithm using RStan package [2018] and consider the posterior samples as the benchmark of the MCMC approach. For each experiment of the factor copula models, we generate 1000 samples via the NUTS algorithm with 500 burn-in. The chain quickly converges after a hundred iterations. Figure 3.7, 3.8 and 3.9 compare the standard deviations of the posterior samples using VI and MCMC for different factor copula models. In one-factor model and nested factor model, the standard deviations of the parameters θ are similar using both methods. In bi-factor copula models, we clearly observe that VI underestimates the standard deviations which is corresponding with the literature of VI estimation, see Blei et al. [2017]. However, the underestimation is acceptable for the copula parameters θ.

Table 3.6 compares the time of computation between VI and the MCMC for 1000 samples. As expected, the computational time of variational approximation is much less than that of the MCMC approach. While the VI convergence time depends on the optimization parameters such as the number of MC samples, the number of MC for calculating the gradients, tolerance, among others. The NUTS depends mainly on the number of iterations. Despite that it is difficult to compare the efficiency of VI and MCMC samples, we obtain a quite reasonable estimate in the limited time with the VI approach. Especially, the complexity of the Student-t copula function makes it really computational expensive for the MCMC approach.

Table 3.6: Time of estimation using VI and NUTS

Copula type	Gaussian	Student-t	Clayton	Gumbel	Frank	Joe	Mix
(a) Time estimated (s) using VI							
One-factor	10	509	24	41	11	17	98
Nested factor	20	783	25	43	13	19	121
Bi-factor	90	3366	165	267	87	154	596
(b) Time estimated (s) using NUTS							
One-factor	1097	491166	3762	5262	1083	2221	8535
Nested factor	1786	567862	5977	6098	1264	3478	20520
Bi-factor	17935	1758007	83680	177932	30345	236809	118729

We report the time of estimation for the factor copula models using VI and NUTS for 1000 samples. The VI convergence time depends on its optimization parameters such as the number of MC samples, number of MC for calculating the gradients, tolerance, among others. The NUTS approach depends mainly on the number of iterations.

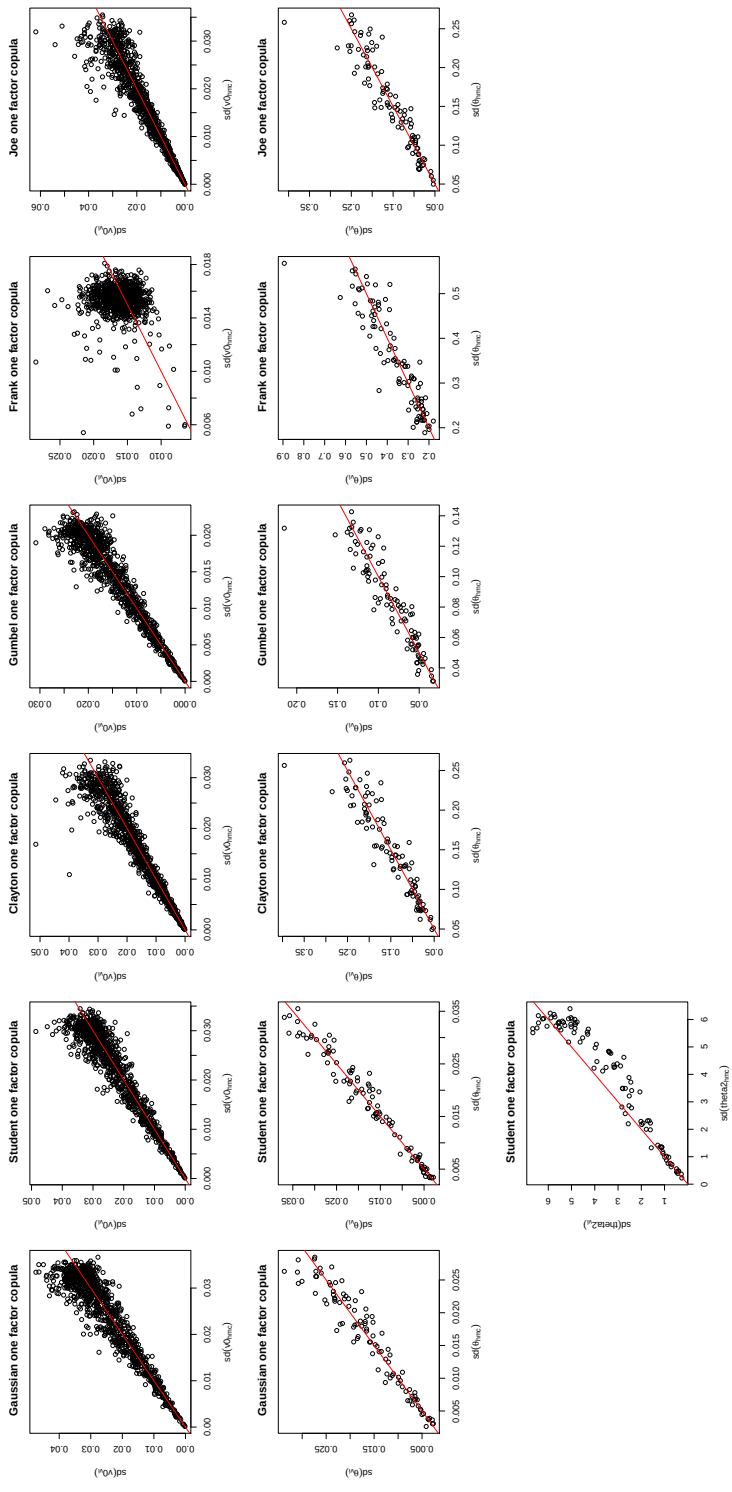

Figure 3.7: Comparison the standard deviations of VI and NUTS estimation for the one-factor copula models.

The figure compares the standard deviations of VI and NUTS estimation for one-factor copula models. In the one-factor models, the standard deviations of the parameters θ are similar using both methods. The VI estimate of the degree of freedom in the factor Student-t copula has a slightly lower variance than that of the MCMC approach.

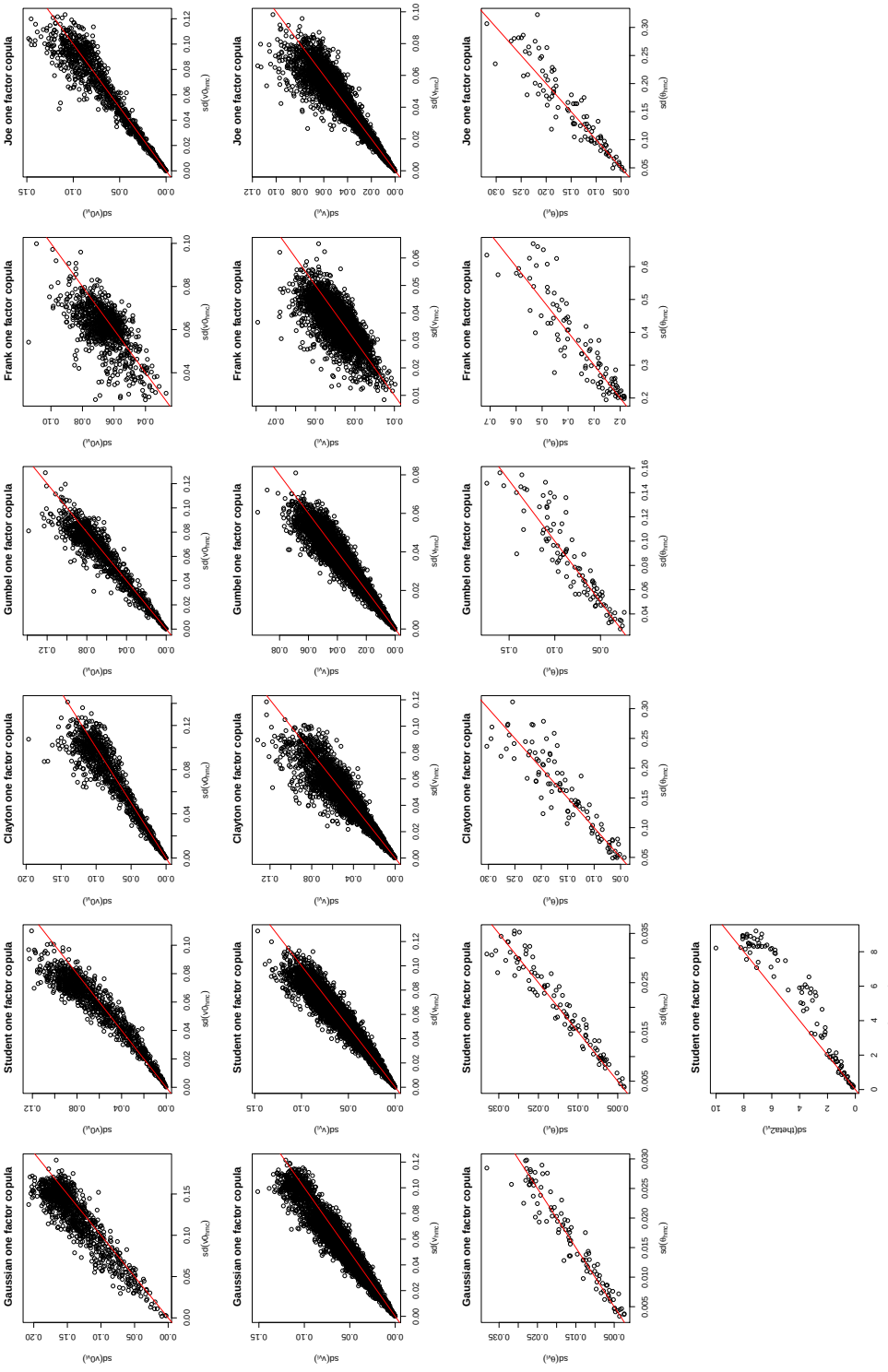

Figure 3.8: Comparison the standard deviations of VI and NUTS estimation for the nested factor copula models.

The figure compares the standard deviation of VI and NUTS estimation for nested-factor copula models. In the nested-factor model, the standard deviations of the parameters θ are similar using both methods. The VI estimate of the degree of freedom in the factor Student-t copula has a slightly lower variance than that of the MCMC approach.

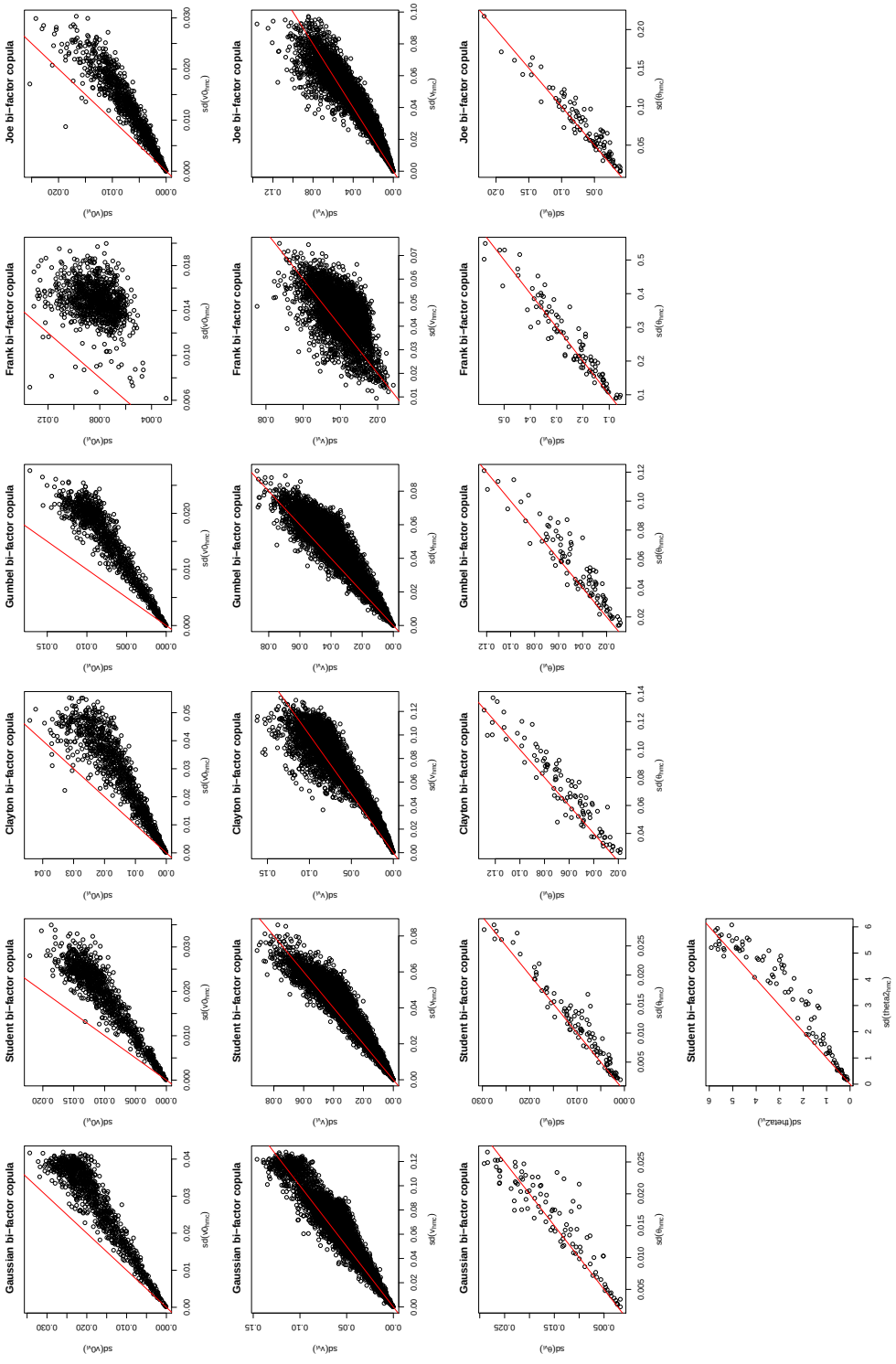

Figure 3.9: Comparison the standard deviations of VI and NUTS estimation for the bi-factor copula models.

The figure compares the standard deviations of VI and NUTS estimation for bi-factor copula models. In the bi-factor model, the standard deviations of the parameters θ and v are lower than that of the MCMC approach. It is acceptable because we are more interested in the copula parameters θ.

3.4 Empirical illustration

In this section, we model the dependence of temperatures measured at 479 stations in Germany. The data is collected from the German Meteorological Service. The time period is selected from 01/01/2013 to 31/12/2015 which results in $T = 1094$ observations. In order to divide stations into groups, we apply a hierarchical clustering approach based on the geographic distance, see Murtagh and Contreras [2011]. We obtain 24 station groups where the distance among stations in each group at most 200 kilometers.

We follow the two-stage estimation procedure for estimating the copula parameters, see Joe [2005], and Chen and Fan [2006]. In the first stage, we find appropriate models for marginal time series and the second stage, we obtain the copula data using the cdf function, $u = (u_{t1}, \dots, u_{td}) = (F_1(x_{t1}), \dots, F_d(x_{td}))$. Then, we apply the selection procedure for one-factor, two-factor, nested factor and bi-factor copula models. The bi-factor model is preferred in both examples when we compare the statistical criteria.

Erhardt et al. [2015] model the marginal distribution of temperatures using the ARMA(1,1) process with the skew Student-t distribution of Fernández and Steel [1998] for the noise,

$$x_{ti} = \alpha_{0i} + \sum_{k=1}^{K} \left(\alpha_{ki} sin\left(\frac{2\pi kt}{365.25}\right) + \beta_{ki} cos\left(\frac{2\pi kt}{365.25}\right) \right) + \phi_i x_{t-1,i} + \epsilon_{ti} + \delta_i \epsilon_{t-1,i},$$

$$\epsilon_{ti} \sim F_{Skew-t}(\nu_i, \gamma_i, \sigma_i),$$

where $(\alpha_{0i}, \phi_i, \delta_i)$ are parameters of ARMA(1,1) process, (ν_i, γ_i) are the parameters of skew Student-t distribution, and $(\alpha_{ki}, \beta_{ki})$ are the slopes of Fourier exogenous regressors with different frequencies to account for the seasonal effect. We choose the value of $K = 2$ to minimize the AIC of marginal models. We estimate the ARMA(1,1) model using MLE and obtain $\hat{\vartheta}_i = \{\hat{\alpha}_{0i}, \hat{\phi}_i, \hat{\delta}_i, \hat{\alpha}_{1i}, \hat{\beta}_{1i}, \hat{\alpha}_{2i}, \hat{\beta}_{2i}, \hat{\nu}_i, \hat{\gamma}_i, \hat{\sigma}_i\}$. We check if the choice of a skew Student-t distribution is suitable for the data by performing the Kolmogorov-Smirnov, Anderson-Darling, and Neyman's smooth tests for the goodness of fit. All series passed the tests with p-values greater than 0.05.

The estimates of ARMA(1,1) process for each marginal series are summarized in the Online Appendix. Then, we obtain the copula data using the cdf transformation of the residuals, $u_{ti} = F_{Skew-t}(\epsilon_{ti}|\hat{\nu}_i, \hat{\gamma}_i, \hat{\sigma}_i)$. Table 3.7 shows the summary statistics for one-factor, two-factor, nested factor and bi-factor copulas using the bivariate copula selection procedure. There are such a large proportion of Student-t copulas that all models reveal strong upper and lower tail dependence between the observable and the common factor. This results in a strong tail dependence of temperatures among stations. The second layer of the two-factor copula and bi-factor copula aim to capture the asymmetric tail dependence and the remain correlation. The relationship among the groups in nested factor copula also confirms this asymmetric dependence. These findings are similar to Erhardt et al. [2015] using a truncated vine copula for a similar problem. In general, the structured factor copula models are more preferred than the multi-factor models. We find that the bi-factor copula model is the most suitable model for the dependence of daily temperatures based on statistical criteria.

Table 3.7: Model comparison of daily temperature dependence

Structure	One-factor	Two-factor	Nested factor	Bi-factor
AIC	-458.9	-611.4	-935.7	-961.5
BIC	-455.4	-606.2	-931.4	-955.3
$\log p(u\|\theta)$	230.1	306.7	468.7	482.0
# bivariate links	479	929	503	958
# Gaussian	1	204	26	365
# Student-t	286	210	442	389
# Clayton (rotated)	0	6	0	1
# Gumbel (rotated)	192	332	35	154
# Frank (rotated)	0	172	0	47
# Joe (rotated)	0	5	0	2
# Independence	0	29	0	0

Next, we can take advantage of the bi-factor copula model to predict the temperatures at different stations from 01/01/2016 to 31/12/2017. We assume that the temperatures are missing at 24 stations during that period, one station for each group sector. Using the information from the 455 remained stations, we want to infer about the temperatures at the missing locations. The idea is first to make inferences on the latent variables during the prediction period, then we

predict the missing temperatures given the latent variable samples. We employ VI to infer the common latent variable and group latent variables. Next, we sample the missing copula variable $u_{ti} \sim F(u_{ti}|v_{t0}, v_{tg})$ and the predicted temperature x_{ti} using the marginal models as,

$$
\begin{aligned}
u_{ti}^{(n)} &\sim F(u_{ti}|v_{t0}^{(n)}, v_{tg}^{(n)}), \\
\epsilon_{ti}^{(n)} &= F_{Skew-t}^{-1}(u_{ti}^{(n)}; \nu_i, \gamma_i, \sigma_i), \\
x_{ti}^{(n)} &= \hat{\alpha}_{0i} + \sum_{k=1}^{K} \left(\hat{\alpha}_{ki} sin\left(\frac{2\pi kt}{365.25}\right) + \hat{\beta}_{ki} cos\left(\frac{2\pi kt}{365.25}\right) \right) + \hat{\phi}_i x_{t-1,i}^{(n)} + \epsilon_{ti}^{(n)} + \hat{\delta}_i \epsilon_{t-1,i}^{(n)},
\end{aligned}
$$

where $v_{t0}^{(n)}, v_{tg}^{(n)}$ are obtained from the variational samples of the predicted latent variables. Note that, each predictive sample $x_{ti}^{(n)}$ requires the previous sample $x_{t-1,i}^{(n)}$, hence the uncertainty is accumulated along the prediction period. Figure 3.10 shows the prediction of temperatures at Grafenberg-Kasberg and Bertsdorf-Hörnitz stations. They are chosen such that in Grafenberg-Kasberg group, we have the information of the 34 other stations while in Bertsdorf-Hörnitz group, we only have the information of the 7 other stations. In general, the predicted mean temperatures together with the 95% prediction interval are quite accurate. We have a better estimation in Grafenberg-Kasberg group because we have more certain estimations of the latent variable in the larger group. The empirical temperatures are included in the prediction interval most of the times.

3.5 Conclusion

In this chapter, we employ the ADVI algorithm in Kucukelbir et al. [2017] to estimate the multi-factor and structured factor copula models. Estimation have been be carried out for high dimensional variables. The posterior means of VI samples are similar to that of MCMC samples while the posterior standard deviations are only underestimated in the case of bi-factor copulas. Instead, VI is much less computational expensive than the MCMC approach. Due to the fast estimation, we take advantage of the posterior modes of the latent variables to reassess the assumption of bivariate copula links and select better bivariate copula functions based on

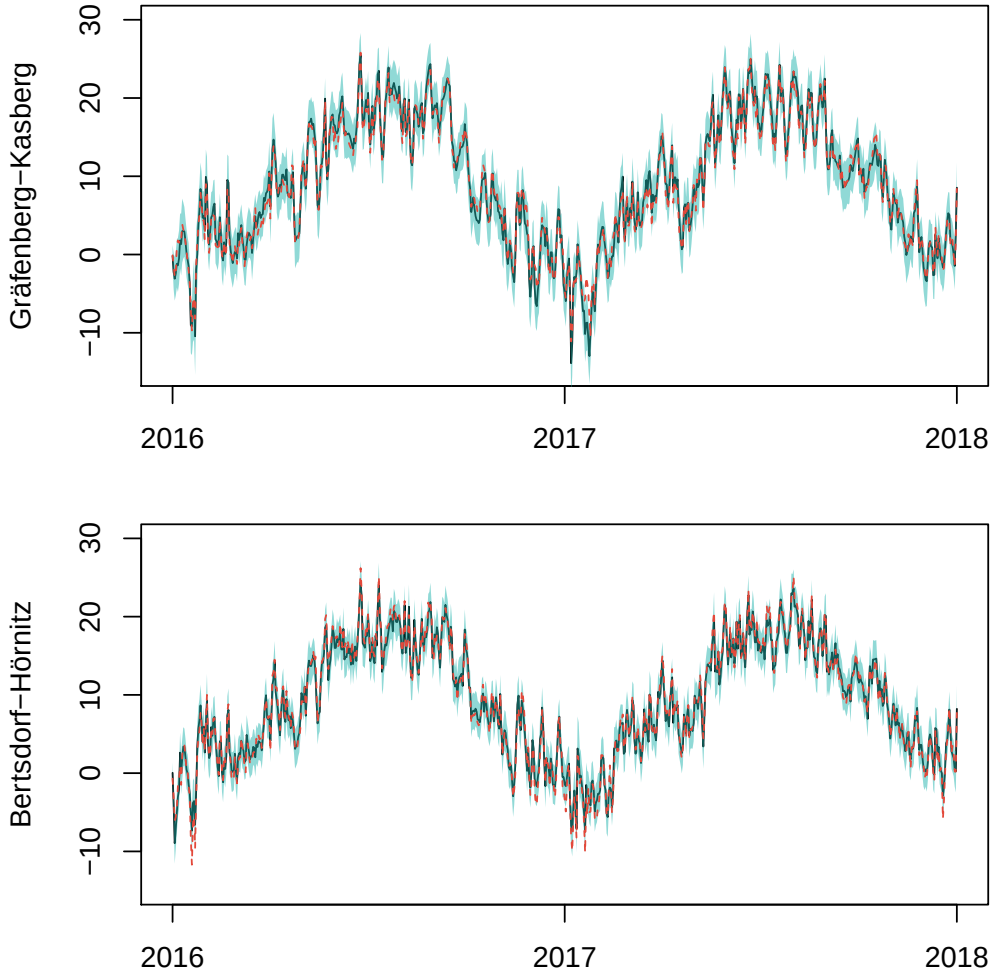

Figure 3.10: The prediction of temperatures using the estimated bi-factor copula model

The figure shows prediction of temperatures using the estimated bi-factor copula model at Grafenberg-Kasberg and Bertsdorf-Hörnitz stations. They are chosen such that in Grafenberg-Kasberg group, we have the information of 34 other stations while in Bertsdorf-Hörnitz group, we only have the information of 7 other stations. The predicted mean temperatures is the black line, and the measured temperatures is the red dash line. The online version of this figure is in color.

minimizing the BIC. The selection procedure performs quite well with different simulated data. In the future, we can allow for switching variables among the group sectors in the structured copula models or extend the static structured factor model to a dynamic factor copula model as proposed in Nguyen et al. [2019].

Chapter 4

Truncated factor vine copula models

In this chapter, we extend the factor copula model proposed by Krupskii and Joe [2013] for a combination of a factor copula at the first tree and truncated vine copulas at higher trees. The new proposed model aims to capture different behaviors at the tail of the joint distribution and provides interpretable economic meanings. This idea is similar to Brechmann and Joe [2014] who construct Gaussian copulas using one factor and one level Gaussian truncated vine, however their proposal ignores the issues of fat tails and asymmetric dependence among variables. Using the VI approach in Nguyen et al. [2018], we estimate the parameters of the truncated factor vine copula models and perform a similar copula selection procedure.

The truncated factor vine copula models promise to outperform the multi-factor copula models due to the flexibility in the tails of the joint distribution. For example, due to different business interactions, groups of financial stock returns usually behave very similarly even conditional on the common market variable. If we account for this dependence by another group factor variable, a small number of series in each group and weak remaining dependence among them results in uncertain estimation of the latent factor and bivariate copula parameters. However, when the number of series in each group increases, their remaining co-movements in extreme events reduce to zero. Therefore, truncated vine copulas in higher tree levels help to capture this conditional dependence more effectively. We illustrate the performance of different factor models for high

dimensional financial returns of 218 blue chip companies listed in 10 different European countries. The truncated factor vine copula model is preferred over the structured copula models with similar level of complexities.

The rest of the chapter is organized as follows. Section 4.1 introduces the proposed truncated factor vine copula models. Section 4.2 presents the Bayesian approach. Section 4.3 shows some numerical simulations of the truncated factor vine copula models. Section 4.4 illustrates the applications with real data. Finally, conclusions are reached in Section 4.5.

4.1 Model specification

Let $X = (X_1, \ldots, X_d)'$ be the d-dimensional continuous random variable and let $F_1(x_1), \ldots, F_d(x_d)$ be their marginal cumulative distribution functions (cdf). Using Sklar [1959] theorem, we can write down the joint cdf function, F, as a copula function of the transformed variables in the unit domain, $F(x_1, \ldots, x_d) = C(u_1, \ldots, u_d)$, where $u_i = F_i(x_i) \in [0, 1]$, for $i = 1, \ldots, d$. Joe [1997] illustrates a large number of bivariate copula functions that suit with different types of joint distributions. However, it is not trivial to extend bivariate copula functions in higher dimensions. In order to capture the co-movements of a large number of variables, regular vine copulas by Bedford and Cooke [2001, 2002] and factor copulas by Krupskii and Joe [2013, 2015a] are commonly used. Taking the good properties of vine copulas and factor copulas, we propose a truncated factor vine copula model that combines a factor copula model at the first tree with a truncated vine copula in higher trees. The factor copula at the first tree accounts for the main relationship among variables and solve for the curse of dimensionality. Then, the remaining dependence among variables are justified by truncated vine copulas.

4.1.1 Truncated factor vine copulas

Truncated factor vine copulas are built based on the pair-copula construction strategy where the copula function density composes of a product of bivariate copula functions, see Aas et al. [2009].

Following the notations in Bedford and Cooke [2001, 2002], Kurowicka and Cooke [2006], and Brechmann et al. [2012], we specify the truncated factor vine copula as a sequence of trees and the nodes in each tree are linked by bivariate copula functions. Let $T_0, T_1, \dots T_K$ be the trees in the truncated factor vine copula model where $K \leq d-1$, and let N_i and E_i be the set of nodes and edges in tree T_i, for $i = 0, \dots, K$ such that the following requirements are satisfied:

(i) Tree T_0 has nodes $N_0 = \{0, 1, \dots, d\}$ and edges $E_0 = \{\{0,1\}, \{0,2\}, \dots, \{0,d\}\}$.

(ii) For $i = 1, \dots, K$, the nodes in tree T_i are the edges in tree T_{i-1}, or $N_i = E_{i-1}$

(iii) If two edges in tree T_i are joined by an edge in tree T_{i+1}, they must share a common node.

The truncated factor vine copula is a specification of regular vine copulas (R-vine) where at tree T_0, there is a latent node staying at the root that links to all observable variables by bivariate copula functions. Let the node set be $N = \{N_0, \dots, N_K\}$ and let the edge set be $E = \{E_0, \dots, E_K\}$, one associates each edge $e(j,k)$ in E_i with a bivariate function $C_{j,k|D(e)}$ and a bivariate copula density $c_{j,k|D(e)}$. The nodes j and k are conditioned nodes and $D(e)$ is the conditioning set. The bivariate copulas in tree T_0 have an empty conditioning set. Let the latent variable at root of tree T_0 be V_0 and let the observable variables be $U = \{U_1, \dots, U_d\}$ where $U_i = F_i(X_i) \sim U(0,1)$. It is straightforward to derive the joint density of the truncated factor vine copula model, see Kurowicka and Cooke [2006], Kurowicka and Cooke [2006], Brechmann et al. [2012], Krupskii and Joe [2013].

$$f(v_0, u_1, \dots, u_d) = \prod_{i=0}^{K} \prod_{e(j,k) \in E_i} c_{j,k|D(e)}(u_{j|D(e)}, u_{k|D(e)}) \tag{4.1}$$

where $u_{j|D(e)} = F(u_j|u_{D(e)})$, $u_{k|D(e)} = F(u_k|u_{D(e)})$ and $u_{D(e)}$ is a sub-vector of u with the indexes $D(e)$. The marginal conditional distribution $F(u_j|u_{D(e)})$ can be computed recursively using the formular derived in Aas et al. [2009]. In general,

$$F(u_j|u_{D(e)}) = \frac{\partial C_{j,i|D(e)^{(-i)}}(F(u_j|u_{D(e)^{(-i)}}), F(u_i|u_{D(e)^{(-i)}}))}{\partial F(u_i|u_{D(e)^{(-i)}})} = F(F(u_j|u_{D(e)^{(-i)}})|F(u_i|u_{D(e)^{(-i)}})) \tag{4.2}$$

where i is a arbitrarily component of $D(e)$ and $D(e)^{(-i)}$ are vector of $D(e)$ excluding i. The truncated factor vine copula model in Brechmann and Joe [2014] is a special case of Eq 4.1 where $c_{j,k|D(e)}$ is a Gaussian copula density. Figure 4.1 shows a combination of a one factor copula model in the first tree and a truncated C-vine copula in the second tree. Figure 4.2 shows a similar structure with a D-vine copula in the second tree. The joint density of a factor-truncated-C-vine model can be simplified as

$$f(v_0, u_1, \ldots, u_d) = \prod_{j=1}^{d} c_{j,0}(u_j, v_0) \prod_{k=2}^{d} c_{1,k|0}(u_{1|0}, u_{k|0}).$$

And the joint density of a factor-truncated-D-vine model can be simplified as

$$f(v_0, u_1, \ldots, u_d) = \prod_{j=1}^{d} c_{j,0}(u_j, v_0) \prod_{k=1}^{d-1} c_{k,k+1|0}(u_{k|0}, u_{k+1|0}).$$

In order to calculate the copula density of observable variables, one can take the integral over the latent space,

$$c(u_1, \ldots, u_d) = \int_0^1 \prod_{i=0}^{K} \prod_{e \in E_i} c_{j,k|D(e)}(u_{j|D(e)}, u_{k|D(e)}) dv_0 \tag{4.3}$$

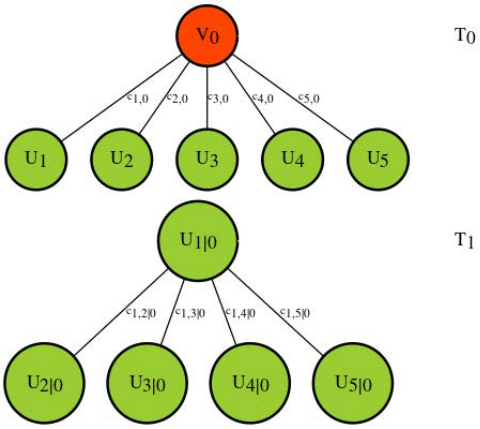

Figure 4.1: An truncated factor vine copula with truncated C-vine for $d = 5, K = 1$

As a truncated factor vine copula model is a special case of a R-vine copula, it is straightforward

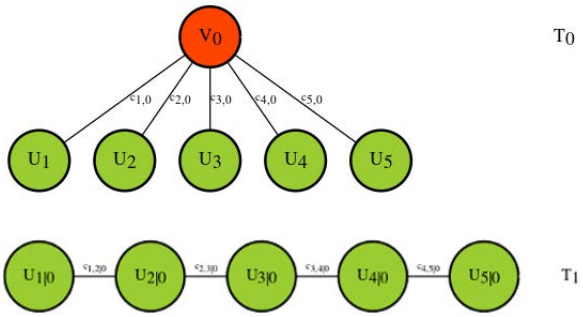

Figure 4.2: An truncated factor vine copula with truncated D-vine for $d = 5, K = 1$

to simulate copula samples using the tree structure. The Algorithm 2.2 in Dissmann et al. [2013] describes the general procedure of sampling a R-vine copula. The idea is to simulate $w_0, \ldots, w_d$ from the uniform distribution $U(0, 1)$ such that

$$\begin{aligned} v_0 &= w_0 \sim U[0,1], \\ u_{1|0} = F_{1|0}(u_1|v_0) &= w_1 \sim U[0,1], \\ &\ldots \\ u_{d|d-1,\ldots,0} = F_{d|d-1,\ldots,0}(u_d|u_{d-1},\ldots,u_1,v_0) &= w_d \sim U[0,1], \end{aligned}$$

Using bivariate copulas and conditional bivariate copulas as building blocks, the marginal conditional distribution function $F(u_i|u_{i-1}, \ldots, u_1, v_0)$ can be obtained using Eq 4.2 and $u_i = F^{-1}(w_i|u_{i-1}, \ldots, u_1, v_0)$.

4.1.2 Discussion

The truncated factor vine copula model can deliver parsimonious and meaningful economic interpretations. For example, stock returns in different countries or different sectors are assumed to be affected by a common economic latent factor. Then, conditional on this latent factor, they also reveal some dependence due to different business interactions which can be captured by truncated vine copulas in higher tree levels. The truncated factor vine copula model can fit well in high dimensional financial time series as it inherits the parsimonious of factor copula models as well

as the flexibility in the tail distribution of vine copula models. In case of hundreds of variables, one can divide variables into groups and fit with smaller vine copula structures in higher trees. However, we still face several challenges of factor copula models as discussed in Nguyen et al. [2018], for example, the latent variable makes it difficult to estimate the factor model even with a known structure. Also, regularly the bivariate links are unknown for most of the empirical datasets. Furthermore, we need to identify the vine structure in each tree and the truncated level in the extended part of the factor model such that the R-vine is parsimonious and still flexible. Kurowicka [2010], Brechmann et al. [2012], and Brechmann and Joe [2015] come up with different strategies to determine the truncated level. Kurowicka [2010] uses a bottom up approach from the highest level and select the tree structure based on the partial correlation matrix obtained from data. On the other hand, Brechmann et al. [2012] start at the level $K = 1$ and extend the tree level one by one until the increase of the goodness of fit is negligible. They compare models using AIC, BIC and the likelihood ratio-based test proposed by Vuong [1989]. While Brechmann and Joe [2015] shows that the procedure can be improved using fit indices. We analyze these issues of the truncated factor vine copula in the next sections.

4.2 Bayesian Inference

In this section, the variational approach is employed to make inferences on both latent variables and bivariate copula parameters. The VI searches for a simpler distribution that comes close to the true posterior distribution in term of minimizing Kullback-Leibler divergence between them. As there are different restrictions in the domains of copula parameters, we transform the constrained domains to the real domain, and approximate the transformed posterior with a product of univariate Gaussian distributions. Then, the stochastic gradient descent is used to find out the free parameters of the proposal Gaussian distributions. For the issue of unknown bivariate copulas in the truncated factor vine copulas, we follow an algorithm in Nguyen et al. [2018] to select the most appropriate functions based on the agreement of the posterior mode of the latent factor and

observable variables. We apply the two-stage estimation procedure assuming that appropriate models for marginal distributions have been found in the first stage, see Joe [2005], and Chen and Fan [2006]. The VI for the parameters of the truncated factor vine copula model is described in the second stage.

4.2.1 Prior distribution

Let $v_0 = \{v_{t0}\}$ denote the common latent factor with the index corresponding to t-th observations, $t = 1, \ldots, T$. Let θ denote the parameters of the bivariate copula functions. Table 3.1 shows the induced prior distributions of θ assuming that the Kendall-τ of bivariate copulas follow an $\mathbb{U}(0, 1)$ distribution for positive dependence copulas and an $\mathbb{U}(-1, 0)$ distribution for negative dependence copulas. Nguyen et al. [2018] discuss the specification of positive and negative dependence copulas in order to tackle the identification problem of factor copula models. We also assume a vague but proper prior distribution $v_0 \sim \mathbb{U}(0, 1)$. And let $\Theta = \{v_0, \theta\}$ be the parameter set of the truncated factor vine copula model.

4.2.2 Posterior distribution

Let $x = (x_{t1}, \ldots, x_{td})'$ be the original data for $t = 1, \ldots, T$ and let $u = (F_1(x_{t1}), \ldots, F_d(x_{td}))'$ be the copula data after fitting an appropriate marginal distribution for each variable. With a specified truncated factor vine copula structure and bivariate links, it is easy to derive the augmented likelihood,

$$p(u|\Theta) = \prod_{t=1}^{T} p(u_{t1}, \ldots, u_{td}|v_{t0}; \theta).$$

Then, the joint posterior density up to a normalized constant is obtained by substituting Eq 4.1,

$$\begin{aligned} p(\Theta|u) \propto p(u, \Theta) &= \prod_{t=1}^{T} p(v_{t0}, u_{t1}, \ldots, u_{td}|\theta)\pi(\theta) \\ &= \prod_{t=1}^{T} \left[\prod_{i=0}^{K} \prod_{e(j,k)\in E_i} c_{j,k|D(e)}(u_{tj|D(e)}, u_{tk|D(e)}) \right] \pi(\theta). \end{aligned} \tag{4.4}$$

If the structure is a factor-truncated-C-vine model as describe before, the posterior density is written as

$$p(\Theta|u) \propto \prod_{t=1}^{T} \left[\prod_{j=1}^{d} c_{j,0}(u_{tj}, v_{t0}) \prod_{k=2}^{d} c_{1,k|0}(u_{t1|0}, u_{tk|0}) \right] \pi(\theta).$$

In case of a factor-truncated-D-vine model, we have

$$p(\Theta|u) \propto \prod_{t=1}^{T} \left[\prod_{j=1}^{d} c_{j,0}(u_{tj}, v_{t0}) \prod_{k=1}^{d-1} c_{k,k+1|0}(u_{tk|0}, u_{tk+1|0}) \right] \pi(\theta).$$

Similar equations can be found for other types of factor-R-vine copulas. In any case, the posterior requires the transformations of copula variables, e.g., $u_{ti|0} = F(u_{ti}|v_{t0})$ whenever we sample v_{t0} or θ, hence the full Bayesian approach is very computational expensive. Instead, we approximate the posterior by a simpler distribution using the Variational approach.

4.2.3 Variational Inference

Given the true posterior of the model is $p(\Theta|u)$ in Eq 4.4, the VI seeks for a proposal density $q(\Theta;\lambda)$, parameterized by a vector λ such that the Kullback–Leibler divergence between them is minimized. However, the posterior $p(\Theta|u)$ is only defined up to a normalizing constant, KL divergence cannot be calculated directly. Instead, the evidence lower bound (ELBO) is used as an equivalent objective function,

$$\begin{aligned} \arg\max_{\lambda} \texttt{ELBO}(q) &= \mathbb{E}_q[\texttt{log}p(u,\Theta)] - \mathbb{E}_q[\texttt{log}q(\Theta;\lambda)] \\ &= \texttt{log}p(u) - KL(q(\Theta;\lambda)||p(\Theta|u)) \leq \texttt{log}p(u) \end{aligned} \tag{4.5}$$

such that if $q(\Theta;\lambda) = p(\Theta|u)$, $\texttt{ELBO} = \texttt{log}p(u)$.

The choice of a proposal density $q(\Theta;\lambda)$ needs to be simple to reduce the computational burden, so practically $q(\Theta;\lambda)$ is written as a product of univariate Gaussian distributions. However, the expectation $\mathbb{E}_q[\texttt{log}p(u,\Theta)]$ cannot be derived in closed form. Instead, the Monte Carlo method can be employed to evaluate the ELBO and then stochastic gradient descent is applied for optimization. In order to reduce the variance of the noisy estimate caused by the Monte Carlo approximation,

we take advantages of the approach by Kucukelbir et al. [2017] based on the reparameterization of the model parameters. The model parameters are transformed from the constrained domains to the real domain, and the transformed posterior is approximated by $q(\Theta; \lambda)$. Also, the learning rate of the stochastic gradient descent decays according to a modified RMSPROP sequence in Tieleman and Hinton [2012]. A detail implementation of the algorithm can be found in Nguyen et al. [2018].

4.2.4 Model check

As discussed in the above, one issue of the truncated factor vine copula is that the bivariate copula links are unknown for high dimensional datasets. We follow the proposed algorithm in Nguyen et al. [2018] to recover these bivariate functions. Firstly, we estimate a specified truncated factor vine copula with random initial bivariate copula functions. Then, we obtain the posterior mode of the common latent factor using the VI, which is fast to make inferences. We consider this posterior mode, $\bar{v}_0$, as the true value of the latent variables if the model specification is correct. Hence, we reassess the agreement of the bivariate copula function between u_i and $\bar{v}_0$ as well as conditional bivariate copulas among $u_{i|0}$ using the BIC. For example, if the original assumption of a bivariate link is a Gaussian copula, but a Student-t copula is more appropriate based on the BIC, we will adjust the original link to the Student-t copula accordingly. If there is any change in the dependence structure, we assign and estimate the new model. By keep doing this until convergence, Nguyen et al. [2018] show that the model goodness of fit increases. The simulation study in the next section also confirms that this is a valid procedure.

To identify the vine structure in each higher tree, we calculate the empirical Kendall-τ matrix of the conditional variables $u_{i|0}$, then we select the spaning tree that maximizes the sum of absolute empirical Kendall-τ as a dependence structure at a higher tree level, see Dissmann et al. [2013]. Then, bivariate copula functions can be specified over the spaning tree. This procedure is the most commonly used in practical applications, see Aas [2016]. Note that, after specifying the dependence structure, the VI approach allows for jointly inferring the common latent variable and the copula parameters rather than estimating parameters tree by tree as suggested by Haff [2012].

In order to determine the level of truncation, one can estimate the likelihood of copula density in Eq 4.3 with different truncation levels and calculate AIC, BIC, likelihood ratio-based test. Alternatively, one can estimate the model at truncation level K and assume that the common latent factor is fixed at the posterior mode and check the independence assumption of the conditional variables at level $K + 1$. For application with financial data, tail dependence can be sufficiently captured with two or three tree levels as the bivariate copula functions in higher trees reveal no tail behavior. Hence, in this case, we use one factor copula combined with one level truncation vine copula in the empirical section.

4.3 Numerical simulation

In this section, VI for truncated factor vine copula models are illustrated through the Monte Carlo study. We generate samples of $d = 100$ variables with $T = 1000$ observation points. We simulate the truncated factor vine copula models with different bivariate copula functions such as all Gaussian copulas, all Student-t copulas, all Clayton copulas, all Frank copulas, all Joe copulas, and mixing among different bivariate copulas. The Kendall's τ correlations among nodes are randomized uniformly in the range $[0.2, 0.8]$.

For each simulated sample, we first use VI to make inferences on the model parameters given their known dependence structure. Then, we use the discussed selection strategy to recover the unknown bivariate links in the truncated factor vine copula model. We also divide the variables into 10 smaller groups in the truncated vines to create more flexibility. We report the time of estimation in seconds using one Intel core i7-4770 @ 3.40GHz processor. The computation can be implemented using the `R` package `vifcopula`.

4.3.1 VI vs MCMC

We compare the time of inference using the VI and the MCMC approach. As the joint posterior distribution can be defined up to a normalizing constant, we apply the No-U-Turn Sampler (NUTS)

as a benchmark of the MCMC approach, see Hoffman and Gelman [2014]. In high dimension, NUTS converges to the target posterior distribution much faster than the Metropolis Hasting or the Gibbs sampling. Therefore, we generate 1000 samples after 500 burn-in samples. Table 4.1 shows the computational time of the VI approximation and MCMC. The time of VI inference mainly depends on the parameters of optimizations such as the number of MC for calculating the `ELBO`, the gradient of `ELBO`, among others. On the other hand, the time of the MCMC depends essentially on the number of samples. Even it is problematic for the comparison, the VI can speed up the inference about 70 times faster than the MCMC approach.

Figure 4.3 compares the posterior means and the posterior standard deviations of model parameters in the truncated factor vine copulas. As shown in the first and the third row for the latent variable and bivariate copula parameters, VI estimates of the posterior means are similar to those using the MCMC. On the second and fourth row, the posterior standard deviations of the common latent variable are similar while VI slightly underestimates the posterior standard deviations of the bivariate copula parameters.

Figure 4.4 shows the contour plot of the posterior samples using VI and MCMC for the mix truncated factor vine copula models. In general, VI approximates quite well the posterior distribution of the latent variable and slightly underestimates the standard deviations of the bivariate copula parameters. The plot also shows that the MCMC posterior samples are not correlated. For that reason, a product of univariate Gaussian distributions can be appropriate as the proposal in the VI.

Table 4.1: Time of estimation using VI and MCMC (in seconds)

Copula type	Gaussian	Student-t	Clayton	Gumbel	Frank	Joe	Mix
VI	80	3282	212	235	70	220	698
MCMC	10269	258714	16602	18984	5938	17919	54083

We report the time of estimation for the factor copula models using VI and MCMC for a sample with $d = 100$ variables and $T = 1000$ observations. The VI convergence time depends mainly on its optimization parameters such as the number of MC samples, number of MC for calculating the gradients, tolerance, among others. The MCMC approach depends essentially on the number of iterations. Here, we obtain 1000 samples from MCMC with 500 burn-in samples.

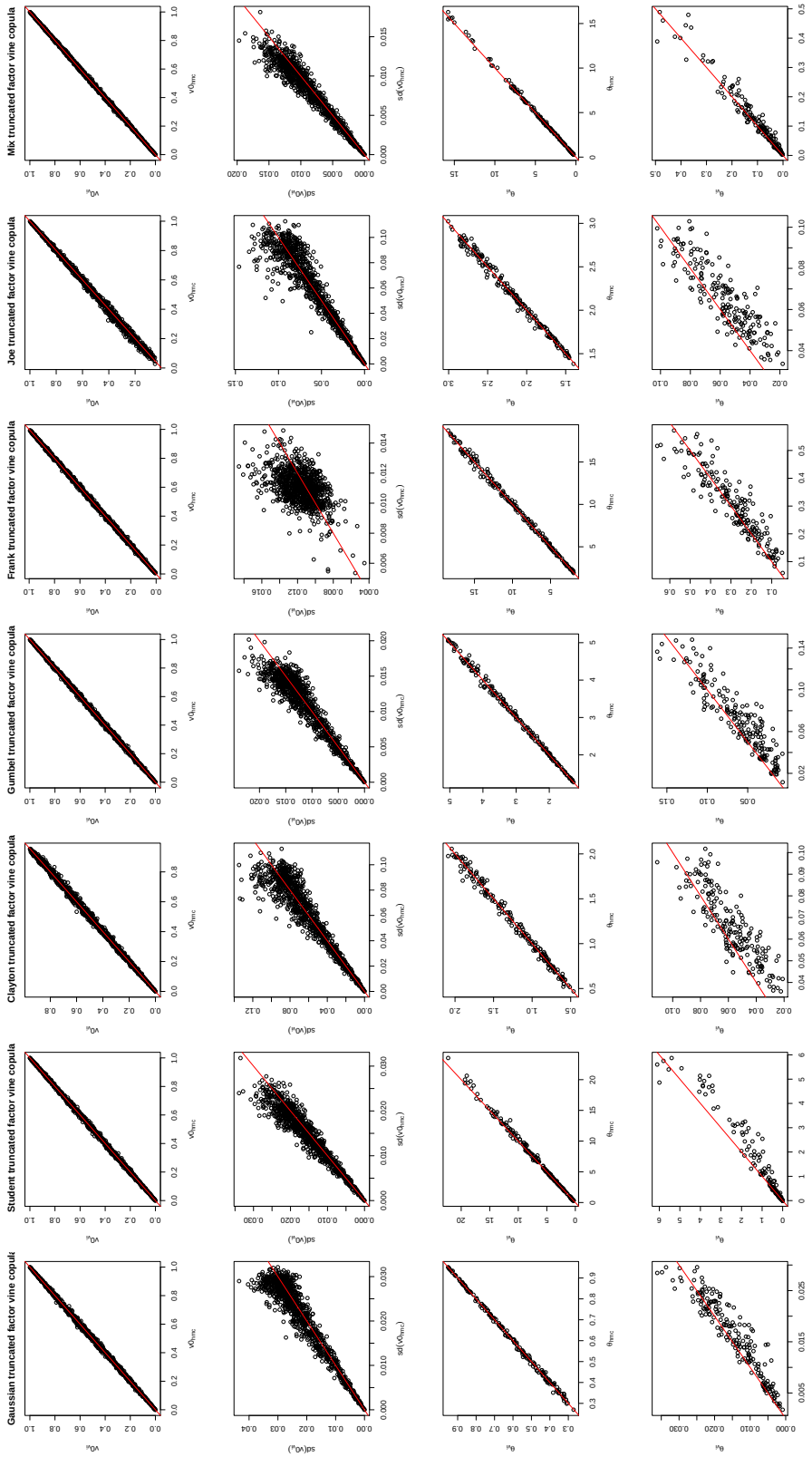

Figure 4.3: Comparison the standard deviations of VI and MCMC estimation for the truncated factor vine copula models.

The figure compares the posterior means and standard deviations of VI and MCMC estimation for the truncated factor vine copula models. The standard deviations of the bivariate copula parameters θ are slightly lower than that of the MCMC approach.

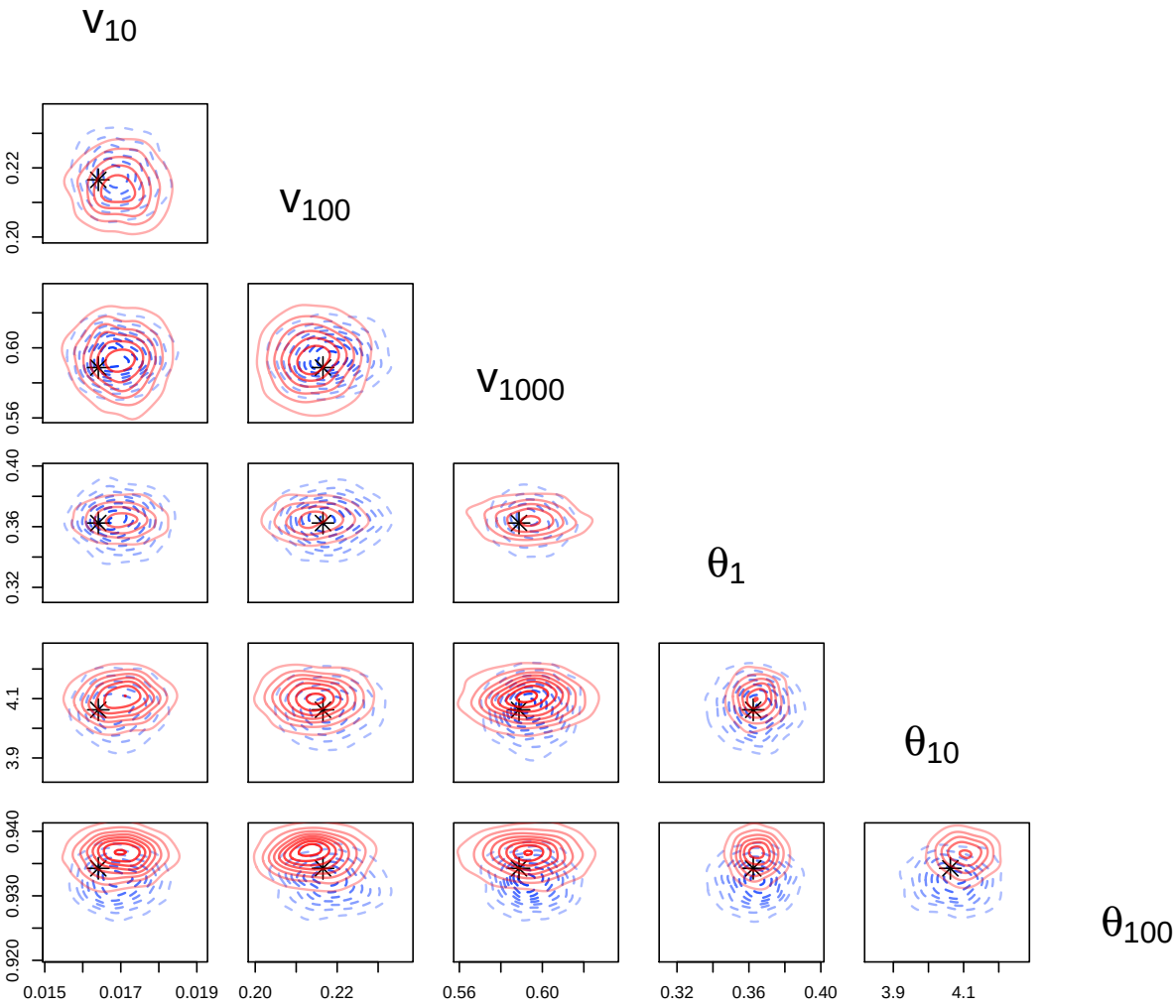

Figure 4.4: The contour plots of posterior samples using VI (red solid lines) and MCMC (blue dashed lines) for the mix truncated factor vine copula models. The true values are marked as black stars.

The figure compares the posterior samples using VI and MCMC for the mix truncated factor vine copula models. In general, VI approximates quite well the posterior distribution of the latent variable and slightly underestimates the standard deviations of the bivariate copula parameters.

4.3.2 Model selection

For a specified truncated factor vine copula model, it is straightforward to use the VI for estimating the model parameters. However, when the bivariate copula functions are unknown, we take advantage of the procedure proposed by Nguyen et al. [2018] to recover this hidden links. Table 4.2 shows the summary statistics and the bivariate copula selection accuracy at tree T_0 and T_1. The selection criteria for specified truncated factor copula models are listed in the Panel (a) and those of unspecified models are presented in Panel (b). The values of ELBO, AIC, BIC, $\log p(u|\theta)$ are normalized for $T = 1000$ observations and they are roughly similar in Panel (a) and Panel (b). Each model has 190 bivariate copula functions with about 190 to 290 bivariate copula parameters. Similar to the issue addressed in Nguyen et al. [2018], the Clayton copula and the survival Joe copula are very alike, so the copula selection often goes wrong which results in lower accuracy in both tree levels. Hence, we restrict the maximum number of selection iterations under 10. Nevertheless, the AIC and BIC even with misspecified models are approximately equal. Secondly, the Student-t truncated factor copula is nearly non-identifiable if both tree level are Student-t copulas, see discussion in Nikoloulopoulos and Joe [2015]. Thus, we use different bivariate copulas in tree T_1 of the Student-t truncated factor copula. As the topology of tree T_0 has been specified, the accuracy in tree T_0 is the percentage that the procedure correctly identifies the original links. Despite that the structure of tree T_1 is not described, using the spanning tree algorithm to maximize the empirical Kendall's τ, the procedure can recover quite accurate the dependence structure in tree T_1. Note that, in this case, we perform the bivariate copula selection for each tree sequentially which results in a slightly smaller values of statistical criteria. If the bivariate copulas are misspecified in tree T_0, they are likely to be misspecified in tree T_1. Performing the jointly selection copulas in all trees are much computationally expensive and it is still an open research question. In general, if the data are generated from a mix truncated factor vine copula, the procedure can recover correctly 85% links in tree T_0 and 62% links in tree T_1

Table 4.2: Model comparison for the truncated factor vine copula models

Copula type	Gaussian	Student-t	Clayton	Gumbel	Frank	Joe	Mix
(a) Initial at the correct structure							
ELBO	60.8	89.2	144.5	130.6	109.1	147.3	113.1
AIC	-122.7	-179.0	-276.4	-259.8	-219.9	-282.6	-217.7
BIC	-121.7	-177.6	-275.5	-258.9	-218.9	-281.6	-216.7
log$p(u\|\theta)$	61.5	89.8	138.4	130.1	110.1	141.5	109.1
(b) Initial at a random structure							
# Selection iteration	4	10	10	6	5	10	10
% accuracy Tree T_0	98	78	71	100	99	64	85
% accuracy Tree T_1	98	86	44	97	95	75	62
ELBO	60.8	88.9	144.2	130.4	108.9	147.1	110.1
AIC	-122.6	-178.5	-276.4	-259.5	-219.4	-282.3	-213.5
BIC	-121.7	-177.2	-275.4	-258.5	-218.4	-281.4	-212.5
log$p(u\|\theta)$	61.5	89.5	138.4	129.9	109.9	141.4	107.0

We report the statistical criteria for the truncated factor vine copula models. Each model has 190 bivariate copula functions with about 190 to 290 bivariate copula parameters. We use Gauss-Legendre quadrature integration over the latent space to obtain log$p(u|\theta)$. The value of ELBO, AIC, BIC, log$p(u|\theta)$ are normalized for 1000 data observations.

4.4 Empirical Illustration

In this section, we estimate the stock return dependence of 218 firms listed in 10 European countries as Austria (12 firms), Belgium (16 firms), Finland (22 firms), France (34 firms), Germany (28 firms), Ireland (10 firms), Italy (31 firms), Netherlands (19 firms), Portugal (15 firms), Spain (31 firms). The selected companies are large blue-chip stocks which are the constituents of stock market indices in 10 European countries. We take data samples from 01/01/2014 to 31/12/2017 which results in $T = 1020$ daily observations. We first filter out the conditional mean and variance of all the marginal stock returns using the AR(1) - EGARCH(1,1) process with the skew Student-t

distribution of Fernández and Steel [1998] for the innovations,

$$
\begin{aligned}
x_{ti} &= \mu_i + \phi_i x_{i,t-1} + \sigma_{it}\epsilon_{it} \\
log(\sigma_{it}^2) &= \alpha_{0i} + \alpha_{1i}\epsilon_{i,t-1} + \delta_i(|\epsilon_{i,t-1}| - E(|\epsilon_{i,t-1}|)) + \beta_i log(\sigma_{i,t-1}^2) \\
\epsilon_{it} &\sim F_{Skew-t}(\nu_i, \gamma_i)
\end{aligned}
$$

where (ν_i, γ_i) are shape parameter and skewness parameter of the skew Student-t distribution, and $(\alpha_{0i}, \alpha_{1i}, \beta_i, \delta_i)$ are the parameters of exponential GARCH model, see Nelson [1991]. We estimate the AR(1) - EGARCH(1,1) model using MLE and obtain the set of marginal parameters $\hat{\vartheta}_i = \{\hat{\mu}_i, \hat{\phi}_i, \hat{\alpha}_{0i}, \hat{\alpha}_{1i}, \hat{\beta}_i, \hat{\delta}_i, \hat{\nu}_i, \hat{\gamma}_i\}$. We also check the goodness of fit of the distribution of residuals ϵ_{it} as in the previous example using Kolmogorov-Smirnov, Anderson-Darling, and Neyman's smooth tests. Every series passes with p-values greater than 0.05. We obtain the copula data using the cdf transformation of the residuals, $u_{it} = F_{Skew-t}(\epsilon_{it}|\nu_i, \gamma_i)$, and model their joint dependence.

Table 4.3 shows the summary statistics of one-factor, two-factor, nested factor, bi-factor, truncated factor vine copula models for stock return data. In the truncated factor vine copula 10 groups, we assume 10 different one level truncated vines for 10 different countries. While the truncated factor vine copula one group assumes no geographical restriction. As expected, we observe a strong tail dependence in the one-factor copula model. The links between observable variables and the common latent variables are mostly Student-t copulas with low degrees of freedom. The two-factor model outperforms the nested factor model which suggests that the joint dependence of European stock returns is only weakly affected by geographic locations. The bi-factor model is still preferable over two-factor model according to the statistical criteria however the improvement is not substantial. The bivariate links in the second tree level of the bi-factor model are mostly Gaussian and Frank copula, thus most of the tail dependence has been captured in the first tree level and the group latent variables do not contribute much to the tail dependence. This is reasonable as after conditional on the market latent factor, the extreme co-movements of a large number of series in each country group reduce to zero. With the same model complexity, we

compare the truncated factor vine copula one group with the two-factor copula. There are more tail dependence bivariate copulas in the truncated factor vine which results in a better model goodness of fit. The truncated factor vine copula 10 groups is also preferred over the bi-factor copula model because of the presence of more heavy tail dependence copulas. The truncated factor vine copula 10 groups is a restricted form of the one group in order to explain the dependence structure among firms based on geographical locations. For example, Figure 4.5 shows the dependence structure of selected firms listed in Austria and Portugal for the truncated factor vine copula 10 groups. However, the dependence structure in the vine tree might change if we add or subtract variables.

Table 4.3: Model comparison of stock return dependence

Structure	One-factor	Two-factor	Nested factor	Bi-factor	Truncated factor vine (1 group)	Truncated factor vine (10 groups)
AIC	-102.4	-117.5	-111.9	-117.7	-134.3	-124.8
BIC	-100.5	-114.6	-109.9	-114.9	-131.3	-121.9
$\texttt{log}p(u\|\theta)$	51.6	59.4	56.3	59.4	67.7	63.0
# bivariate links	218	434	228	412	435	426
# Gaussian	11	19	7	61	61	51
# Student-t	176	171	177	177	176	181
# Clayton (rotated)	0	0	0	4	5	8
# Gumbel (rotated)	20	37	34	51	53	57
# Frank (rotated)	11	207	10	115	140	129
# Joe (rotated)	0	0	0	1	0	0
# Independence	0	2	0	24	0	0

We report the statistical criteria for the different factor copula models. We use Gauss-Legendre quadrature integration over the latent space to obtain $\texttt{log}p(u|\theta)$. The value of ELBO, AIC, BIC, $\texttt{log}p(u|\theta)$ are normalized for 1020 data observations.

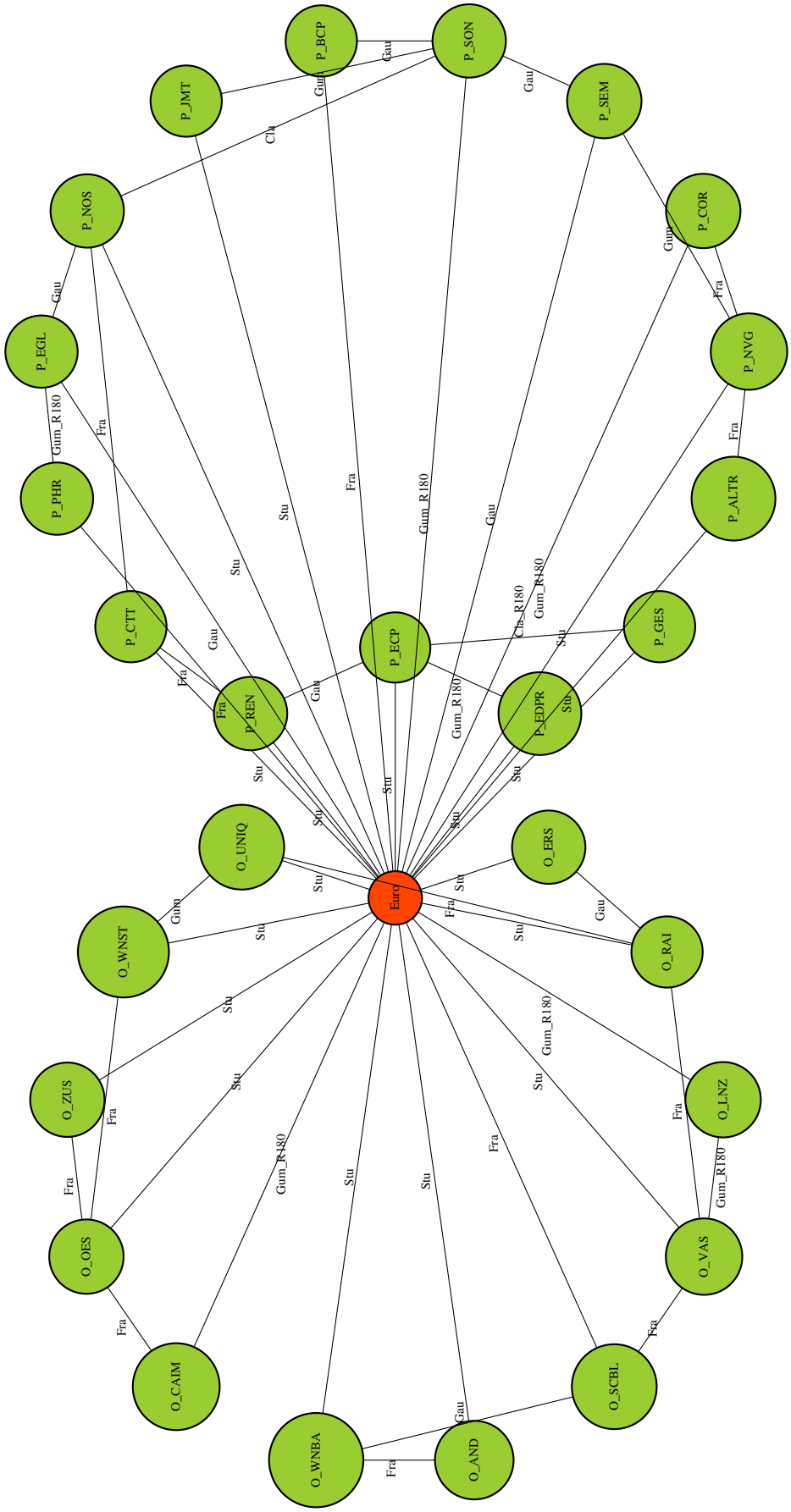

Figure 4.5: Dependence structure of selected firms listed in Austria and Portugal

Next, we compare the Spearman's ρ_S and the tail-weighted dependence measures of the truncated factor vine copula one group to that of the empirical copula data. Krupskii and Joe [2015b] propose the tail-weighted dependence as the correlation of transformed variables where the joint movements in the tails have heavier weights,

$$\begin{aligned}\rho_L &= Cor\left((1-\frac{U_1}{p})^6, (1-\frac{U_2}{p})^6 \middle| U_1 < p, U_2 < p\right)\\ \rho_U &= Cor\left((1-\frac{1-U_1}{p})^6, (1-\frac{1-U_2}{p})^6 \middle| 1-U_1 < p, 1-U_2 < p\right)\end{aligned} \tag{4.6}$$

In order to calculate ρ_S, ρ_L, ρ_U, we simulate pseudo samples from the the truncated factor vine copula one group and take the average of ρ_S, ρ_L, ρ_U as an estimate for the model implied rank correlation and tail dependence.

Figure 4.6 shows the heat maps of the Spearman's ρ matrix in the first row and the lower and upper tail-weighted dependence measures in the last two rows. The figures have a similar pattern and the truncated factor vine copula model can capture well the correlation and the tail dependence. The average of the absolute difference of Spearman's ρ between the truncated factor vine copula model and that of the empirical copula data is 0.036 while the average difference of lower tail-weighted dependence measure is 0.060 and upper tail-weighted dependence measure is 0.070. The results are similar with the measures using bi-factor copulas in Nguyen et al. [2018], where $\rho_S = 0.032, \rho_L = 0.059, \rho_U = 0.073$.

4.5 Conclusion

In this chapter, we have proposed a truncated factor vine copula model as an extension of factor copula models. The truncated factor vine copula model is built based on the pair-copula construction strategy with a combination of a one factor copula in the first tree and truncated vine copulas in the higher tree levels. The model inherits good properties of factor copulas and vine copulas such as the parsimony and flexibility for different tail behaviors. The VI approach helps to make a fast inference on the latent variables that helps to recover the dependence structure of the

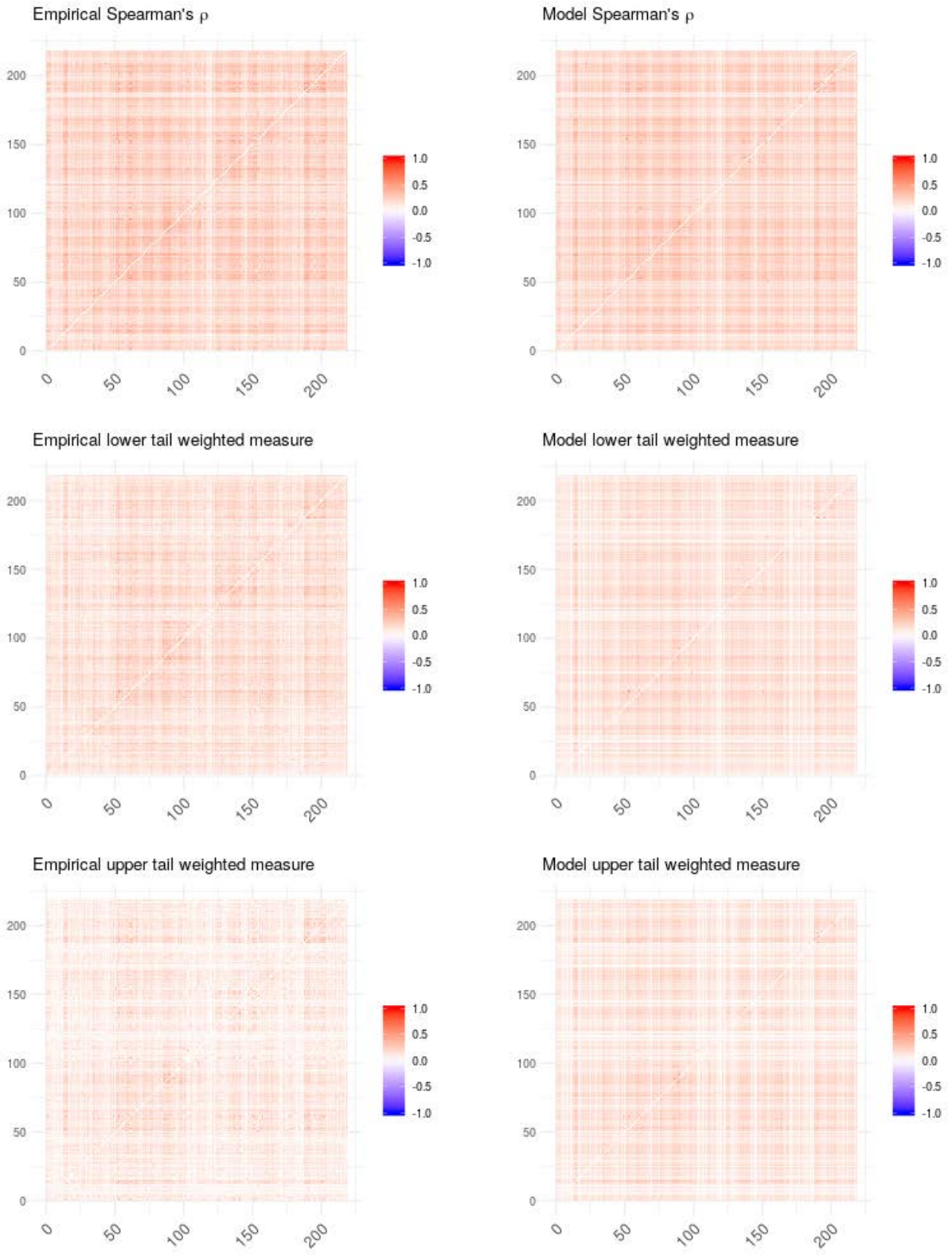

Figure 4.6: The Spearman's ρ and the tail-weighted dependence measures of the empirical copula and the truncated factor vine copula model

models. Compared to the structured factor copula model, the truncated factor vine copula models are preferred with similar structure complexities.

As an extension, we can simplify the vine copulas in higher tree levels by Gaussian bivariate copulas as most of the tail dependence is usually captured by Student-t and asymmetric Archimedean copulas in the first tree level. Besides, one can develop the model by substituting the truncated vine copulas in higher tree levels with multivariate or hierarchical Archimedean copulas for small groups of variables and reduce the number of parameters in higher levels.

References

K. Aas. Pair-copula constructions for financial applications: A review. *Econometrics*, 4(4):43, 2016.

K. Aas and I. H. Haff. The generalized hyperbolic skew students t-distribution. *Journal of Financial Econometrics*, 4(2):275–309, 2006.

K. Aas, C. Czado, A. Frigessi, and H. Bakken. Pair-copula constructions of multiple dependence. *Insurance: Mathematics and Economics*, 44(2):182–198, 2009.

G. J. Alexander and A. M. Baptista. A comparison of VaR and CVaR constraints on portfolio selection with the mean-variance model. *Management Science*, 50(9):1261–1273, 2004.

L. Andersen and J. Sidenius. Extensions to the Gaussian copula: Random recovery and random factor loadings. *Journal of Credit Risk*, 1(1):29–70, 2004.

G. W. Bassett Jr, R. Koenker, and G. Kordas. Pessimistic portfolio allocation and choquet expected utility. *Journal of Financial Econometrics*, 2(4):477–492, 2004.

T. Bedford and R. M. Cooke. Probability density decomposition for conditionally dependent random variables modeled by vines. *Annals of Mathematics and Artificial Intelligence*, 32(1-4): 245–268, 2001.

T. Bedford and R. M. Cooke. Vines: A new graphical model for dependent random variables. *Annals of Statistics*, pages 1031–1068, 2002.

F. Blasques, S. J. Koopman, and A. Lucas. Information-theoretic optimality of observation-driven time series models for continuous responses. *Biometrika*, 102(2):325–343, 2015.

D. M. Blei, A. Kucukelbir, and J. D. McAuliffe. Variational inference: A review for statisticians. *Journal of the American Statistical Association*, 112(518):859–877, 2017.

T. Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31 (3):307–327, 1986.

E. C. Brechmann and H. Joe. Parsimonious parameterization of correlation matrices using truncated vines and factor analysis. *Computational Statistics & Data Analysis*, 77:233–251, 2014.

E. C. Brechmann and H. Joe. Truncation of vine copulas using fit indices. *Journal of Multivariate Analysis*, 138:19–33, 2015.

E. C. Brechmann, C. Czado, and K. Aas. Truncated regular vines in high dimensions with application to financial data. *Canadian Journal of Statistics*, 40(1):68–85, 2012.

H. Chen, R. MacMinn, and T. Sun. Multi-population mortality models: A factor copula approach. *Insurance: Mathematics and Economics*, 63:135–146, 2015.

X. Chen and Y. Fan. Estimation and model selection of semiparametric copula-based multivariate dynamic models under copula misspecification. *Journal of Econometrics*, 135(1-2):125–154, 2006.

U. Cherubini, S. Mulinacci, F. Gobbi, and S. Romagnoli. *Dynamic copula methods in finance*. John Wiley & Sons, 2011.

D. Creal, S. J. Koopman, and A. Lucas. A dynamic multivariate heavy-tailed model for time-varying volatilities and correlations. *Journal of Business & Economic Statistics*, 29(4):552–563, 2011.

D. Creal, S. J. Koopman, and A. Lucas. Generalized autoregressive score models with applications. *Journal of Applied Econometrics*, 28(5):777–795, 2013.

D. D. Creal and R. S. Tsay. High dimensional dynamic stochastic copula models. *Journal of Econometrics*, 189(2):335 – 345, 2015.

T. Datastream. Thomson reuters datastream. *In Constituents of the S&P 500*, (Retrieved: March 2018 from Datastream database).

S. Demarta and A. J. McNeil. The t copula and related copulas. *International Statistical Review*, 73 (1):111–129, 2005.

A. Dias and P. Embrechts. Modeling exchange rate dependence dynamics at different time horizons. *Journal of International Money and Finance*, 29(8):1687–1705, 2010.

J. Dissmann, E. C. Brechmann, C. Czado, and D. Kurowicka. Selecting and estimating regular vine copulae and application to financial returns. *Computational Statistics & Data Analysis*, 59:52–69, 2013.

P. Embrechts. Copulas: A personal view. *Journal of Risk and Insurance*, 76(3):639–650, 2009.

R. Engle. Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business & Economic Statistics*, 20 (3):339–350, 2002.

R. Engle and B. Kelly. Dynamic equicorrelation. *Journal of Business & Economic Statistics*, 30(2): 212–228, 2012.

T. M. Erhardt, C. Czado, and U. Schepsmeier. R-vine models for spatial time series with an application to daily mean temperature. *Biometrics*, 71(2):323–332, 2015.

J. Fan and R. Li. Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association*, 96(456):1348–1360, 2001.

Y. Fan and A. J. Patton. Copulas in econometrics. *Annual Review of Economics*, 6:179–200, 2014.

C. Fernández and M. F. Steel. On Bayesian modeling of fat tails and skewness. *Journal of the American Statistical Association*, 93(441):359–371, 1998.

A. Ghalanos. *rugarch: Univariate GARCH models.*, 2018. R package version 1.4-0.

L. R. Glosten, R. Jagannathan, and D. E. Runkle. On the relation between the expected value and the volatility of the nominal excess return on stocks. *The Journal of Finance*, 48(5):1779–1801, 1993.

I. H. Haff. Comparison of estimators for pair-copula constructions. *Journal of Multivariate Analysis*, 110:91–105, 2012.

I. H. Haff, K. Aas, and A. Frigessi. On the simplified pair-copula construction simply useful or too simplistic. *Journal of Multivariate Analysis*, 101(5):1296–1310, 2010.

A. C. Harvey. *Dynamic models for volatility and heavy tails: with applications to financial and economic time series*, volume 52. Cambridge University Press, 2013.

M. D. Hoffman and A. Gelman. The No-U-turn sampler: adaptively setting path lengths in hamiltonian monte carlo. *Journal of Machine Learning Research*, 15(1):1593–1623, 2014.

J. Hull and A. White. Valuation of a CDO and an n-th to default CDS without monte carlo simulation. *The Journal of Derivatives*, 12(2):8–23, 2004.

R. Jagannathan and T. Ma. Risk reduction in large portfolios: Why imposing the wrong constraints helps. *The Journal of Finance*, 58(4):1651–1683, 2003.

H. Joe. Multivariate extreme-value distributions with applications to environmental data. *Canadian Journal of Statistics*, 22(1):47–64, 1994.

H. Joe. Families of m-variate distributions with given margins and m (m-1)/2 bivariate dependence parameters. *Lecture Notes-Monograph Series*, pages 120–141, 1996.

H. Joe. *Multivariate models and multivariate dependence concepts*, volume 73. CRC Press, 1997.

H. Joe. Asymptotic efficiency of the two-stage estimation method for copula-based models. *Journal of Multivariate Analysis*, 94(2):401–419, 2005.

H. Joe. *Dependence modeling with copulas*. CRC Press, 2014.

H. Joe and D. Kurowicka. *Dependence modeling: vine copula handbook*. World Scientific, 2010.

T.-H. Kim and H. White. On more robust estimation of skewness and kurtosis. *Finance Research Letters*, 1(1):56–73, 2004.

R. Koenker and G. Bassett Jr. Regression quantiles. *Econometrica*, pages 33–50, 1978.

S. J. Koopman, A. Lucas, and M. Scharth. Predicting time-varying parameters with parameter-driven and observation-driven models. *Review of Economics and Statistics*, 98(1): 97–110, 2016.

P. Krupskii and M. G. Genton. Factor copula models for data with spatio-temporal dependence. *Spatial Statistics*, 22:180–195, 2017.

P. Krupskii and H. Joe. Factor copula models for multivariate data. *Journal of Multivariate Analysis*, 120:85 – 101, 2013.

P. Krupskii and H. Joe. Structured factor copula models: Theory, inference and computation. *Journal of Multivariate Analysis*, 138:53–73, 2015a.

P. Krupskii and H. Joe. Tail-weighted measures of dependence. *Journal of Applied Statistics*, 42(3): 614–629, 2015b.

P. Krupskii, R. Huser, and M. G. Genton. Factor copula models for replicated spatial data. *Journal of the American Statistical Association*, 0(ja):0–0, 2016.

A. Kucukelbir, D. Tran, R. Ranganath, A. Gelman, and D. M. Blei. Automatic differentiation variational inference. *The Journal of Machine Learning Research*, 18(1):430–474, 2017.

D. Kurowicka. Optimal truncation of vines. In *Dependence modeling: Vine copula handbook*, pages 233–247. World Scientific, 2010.

D. Kurowicka and R. M. Cooke. *Uncertainty analysis with high dimensional dependence modelling*. John Wiley & Sons, 2006.

D. Lee and H. Joe. Multivariate extreme value copulas with factor and tree dependence structures. *Extremes*, pages 1–30, 2017.

E. R. Lee, H. Noh, and B. U. Park. Model selection via bayesian information criterion for quantile regression models. *Journal of the American Statistical Association*, 109(505):216–229, 2014.

A. Lucas, J. Schaumburg, and B. Schwaab. Bank business models at zero interest rates. *Journal of Business & Economic Statistics*, (to appear), 2018.

H. Manner and J. Segers. Tails of correlation mixtures of elliptical copulas. *Insurance: Mathematics and Economics*, 48(1):153–160, 2011.

H. Markowitz. Portfolio selection. *The Journal of Finance*, 7(1):77–91, 1952.

A. J. McNeil, R. Frey, and P. Embrechts. *Quantitative risk management: concepts, techniques, and tools*. Princeton university press, 2010.

O. Morales-Nápoles. *Counting vines*. World Scientific, 2010.

J. S. Murray, D. B. Dunson, L. Carin, and J. E. Lucas. Bayesian Gaussian copula factor models for mixed data. *Journal of the American Statistical Association*, 108(502):656–665, 2013.

F. Murtagh and P. Contreras. Methods of hierarchical clustering. *arXiv preprint arXiv:1105.0121*, 2011.

D. B. Nelson. Conditional heteroskedasticity in asset returns: A new approach. *Econometrica*, pages 347–370, 1991.

H. Nguyen, M. C. Ausín, and P. Galeano. Variational inference for high dimensional structured factor copulas. *UC3M Working Papers Statistics and Econometrics*, WP18-05, 2018.

H. Nguyen, M. C. Ausín, and P. Galeano. Parallel bayesian inference for high dimensional dynamic factor copulas. *Journal of Financial Econometrics*, 17(1):118–151, 2019.

A. K. Nikoloulopoulos and H. Joe. Factor copula models for item response data. *Psychometrika*, 80 (1):126–150, 2015.

D. H. Oh and A. J. Patton. Modelling dependence in high dimensions with factor copulas. *Journal of Business & Economic Statistics*, 35(1), 2017a.

D. H. Oh and A. J. Patton. Time-varying systemic risk: Evidence from a dynamic copula model of CDS spreads. *Journal of Business & Economic Statistics*, 2017b. doi: 10.1080/07350015.2016.1177535.

O. Okhrin, Y. Okhrin, and W. Schmid. On the structure and estimation of hierarchical archimedean copulas. *Journal of Econometrics*, 173(2):189–204, 2013.

A. J. Patton. Modelling asymmetric exchange rate dependence. *International Economic Review*, 47 (2):527–556, 2006.

A. J. Patton. A review of copula models for economic time series. *Journal of Multivariate Analysis*, 110:4–18, 2012.

G. O. Roberts and J. S. Rosenthal. Examples of adaptive MCMC. *Journal of Computational and Graphical Statistics*, 18(2):349–367, 2009.

RStan package. RStan: the R interface to Stan, 2018. URL http://mc-stan.org/. R package version 2.17.3.

C. Savu and M. Trede. Hierarchies of archimedean copulas. *Quantitative Finance*, 10(3):295–304, 2010.

U. Schepsmeier and J. Stöber. Derivatives and fisher information of bivariate copulas. *Statistical Papers*, 55(2):525–542, 2014.

D. Scott and F. Grimson. Skewhyperbolic: The skew hyperbolic student t-distribution. 2015. URL https://CRAN.R-project.org/package=SkewHyperbolic. R package version 0.3-2.

M. Sklar. *Fonctions de repartition a n dimensions et leurs marges*. Universite Paris 8, 1959.

M. S. Smith. Bayesian approaches to copula modelling. *Available at SSRN 1974297*, 2011.

R. Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society, Series B (Methodological)*, pages 267–288, 1996.

T. Tieleman and G. Hinton. Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude. *COURSERA: Neural networks for machine learning*, 4(2):26–31, 2012.

M. van der Voort. Factor copulas: External defaults. *The Journal of Derivatives*, 14(3):94–102, 2007.

Q. H. Vuong. Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica: Journal of the Econometric Society*, pages 307–333, 1989.

Q. Xu, Y. Zhou, C. Jiang, K. Yu, and X. Niu. A large CVaR-based portfolio selection model with weight constraints. *Economic Modelling*, 59:436–447, 2016.

C. Zhang, J. Butepage, H. Kjellstrom, and S. Mandt. Advances in variational inference. *arXiv preprint arXiv:1711.05597*, 2018.

Appendix A

Appendix of Chapter 2

A.1 Score update for the factor copula model

A.1.1 Dynamic Gaussian one factor copula

The conditional cdf of $u_t = (u_{1t}, \ldots, u_{dt})'$, where $u_{it} = \Phi(x_{it})$, is given by:

$$\begin{aligned} F(u_{1t}, \ldots, u_{dt} \mid z_t, f_t, \mathcal{F}_t, \theta) &= \Pr(U_{1t} \leq u_{1t}, \ldots, U_{dt} \leq u_{dt} \mid z_t, f_t, \mathcal{F}_t, \theta) \\ &= \Pr\left(X_{1t} \leq \Phi^{-1}(u_{1t}), \ldots, X_{dt} \leq \Phi^{-1}(u_{dt}) \mid z_t, f_t, \mathcal{F}_t, \theta\right) \\ &= \prod_{i=1}^{d} \Pr\left(X_{it} \leq \Phi^{-1}(u_{it}) \mid z_t, f_t, \mathcal{F}_t, \theta\right). \end{aligned}$$

Now, note that, given $\{z_t, f_t, \mathcal{F}_t, \theta\}$, the correlation ρ_{it} is known and X_{it} follows a Gaussian distribution with mean $\rho_{it} z_t$ and standard deviation $\sqrt{1-\rho_{it}^2}$. Then, the conditional density of u_t is,

$$p(u_t \mid z_t, f_t, \mathcal{F}_t, \theta) = \frac{\partial^d F(u_{1t}, \ldots, u_{dt} \mid z_t, f_t, \mathcal{F}_t, \theta)}{\partial u_{1t} \ldots \partial u_{dt}} = \prod_{i=1}^{d} \frac{\phi\left(\Phi^{-1}(u_{it}) \mid \rho_{it} z_t, \sqrt{1-\rho_{it}^2}\right)}{\phi(\Phi^{-1}(u_{it}) \mid 0, 1)},$$

where $\phi(\cdot \mid \mu, \sigma)$ denotes a normal pdf with mean, μ, and standard deviation, σ. Then, the proposed dynamic process is based on the derivative of the log conditional density wrt the dynamic f_{it}, i.e.:

$$\begin{aligned}
s_{it} &= \frac{\partial \log p\left(u_t \mid z_t, f_t, \mathcal{F}_t, \theta\right)}{\partial f_{it}} = \frac{\partial \log p\left(u_t \mid z_t, f_t, \mathcal{F}_t, \theta\right)}{\partial \rho_{it}} \frac{\partial \rho_{it}}{\partial f_{it}} \\
&= \frac{\partial \sum_{i=1}^{d} \left(\log \phi\left(\Phi^{-1}\left(u_{it}\right) \mid \rho_{it} z_t, \sqrt{1-\rho_{it}^2}\right) - \log \phi\left(\Phi^{-1}\left(u_{it}\right) \mid 0, 1\right)\right)}{\partial \rho_{it}} \frac{1-\rho_{it}^2}{2} \\
&= \frac{\partial \left(-\frac{1}{2}\log(2\pi) - \frac{1}{2}\log(1-\rho_{it}^2) - \frac{1}{2}\frac{(\Phi^{-1}(u_{it}) - \rho_{it} z_t)^2}{1-\rho_{it}^2}\right)}{\partial \rho_{it}} \frac{1-\rho_{it}^2}{2} \\
&= \left(\frac{\rho_{it}}{(1-\rho_{it}^2)} + \frac{z_t(\Phi^{-1}\left(u_{it}\right) - \rho_{it} z_t)}{1-\rho_{it}^2} - \frac{\rho_{it}(\Phi^{-1}\left(u_{it}\right) - \rho_{it} z_t)^2}{(1-\rho_{it}^2)^2}\right) \frac{1-\rho_{it}^2}{2} \\
&= \frac{1}{2}\Phi^{-1}\left(u_{it}\right) z_t + \frac{1}{2}\rho_{it} - \rho_{it}\frac{\Phi^{-1}\left(u_{it}\right)^2 + z_t^2 - 2\rho_{it}\Phi^{-1}\left(u_{it}\right) z_t}{2(1-\rho_{it}^2)},
\end{aligned}$$

which leads to the expression given in (2.3).

A.1.2 Dynamic generalized hyperbolic skew Student-t one factor copula

The conditional cdf of $u_t = (u_{1t}, \ldots, u_{dt})$, where $u_{it} = F_{GSt}\left(x_{it} \mid \nu, \gamma\right)$, is:

$$\begin{aligned}
F\left(u_{1t}, \ldots, u_{dt} \mid z_t, \zeta_t, f_t, \mathcal{F}_t, \theta\right) &= \Pr\left(X_{1t} \leq F_{GSt}^{-1}\left(u_{1t} \mid \nu, \gamma\right), \ldots, X_{dt} \leq F_{GSt}^{-1}\left(u_{dt} \mid \nu, \gamma\right) \mid z_t, \zeta_t, f_t, \mathcal{F}_t, \theta\right) \\
&= \prod_{i=1}^{d} \Pr\left(\tilde{X}_{it} \leq \frac{F_{GSt}^{-1}\left(u_{it} \mid \nu, \gamma\right) - \gamma\zeta_t}{\sqrt{\zeta_t}} \mid z_t, \zeta_t, f_t, \mathcal{F}_t, \theta\right),
\end{aligned}$$

where $\tilde{X}_{it} = \left(X_{it} - \gamma\zeta_t\right)/\sqrt{\zeta_t}$. Similarly, given $\{z_t, \zeta_t, f_t, \mathcal{F}_t, \theta\}$, the correlation ρ_{it} is known and $\tilde{X}_{it}$ follows a Gaussian distribution with mean $\rho_{it} z_t$ and standard deviation $\sqrt{1-\rho_{it}^2}$. Then, the conditional density of u_t is,

$$p\left(u_t \mid z_t, \zeta_t, f_t, \mathcal{F}_t, \theta\right) = \frac{\partial^d F\left(u_{1t}, \ldots, u_{dt} \mid z_t, \zeta_t, f_t, \mathcal{F}_t, \theta\right)}{\partial u_{1t} \ldots \partial u_{dt}} = \prod_{i=1}^{d} \frac{\phi\left(\frac{F_{GSt}^{-1}(u_{it}|\nu) - \gamma\zeta_t}{\sqrt{\zeta_t}} \mid \rho_{it} z_t, \sqrt{1-\rho_{it}^2}\right)}{f_{GSt}\left(F_{GSt}^{-1}\left(u_{it} \mid \nu, \gamma\right) \mid \nu, \gamma\right)\sqrt{\zeta_t}}, \tag{A.1}$$

where $f_{GSt}(\cdot \mid \nu, \gamma)$ denotes the standard generalized hyperbolic skew Student-t with ν degrees of freedom and γ skewness parameter. Thus, the equation for s_{it} remains,

$$
\begin{aligned}
s_{it} &= \frac{\partial \log p\left(u_t \mid z_t, \zeta_t, f_t, \mathcal{F}_t, \theta\right)}{\partial f_{it}} = \frac{\partial \log \phi\left(\frac{F_{GSt}^{-1}(u_{it}|\nu)-\gamma\zeta_t}{\sqrt{\zeta_t}} \mid \rho_{it} z_t, \sqrt{1-\rho_{it}^2}\right)}{\partial \rho_{it}} \frac{1-\rho_{it}^2}{2} \\
&= \frac{1}{2} \frac{F_{GSt}^{-1}\left(u_{it} \mid \nu\right) - \gamma\zeta_t}{\sqrt{\zeta_t}} z_t + \frac{1}{2}\rho_{it} - \rho_{it} \frac{\left(\frac{F_{GSt}^{-1}(u_{it}|\nu)-\gamma\zeta_t}{\sqrt{\zeta_t}}\right)^2 + z_t^2 - 2\rho_{it}\frac{F_{GSt}^{-1}(u_{it}|\nu)-\gamma\zeta_t}{\sqrt{\zeta_t}} z_t}{2(1-\rho_{it}^2)},
\end{aligned}
$$

which leads to the expression given in (2.5).

A.2 Equivalence of predictive density

Here, we show that our GAS update equation is similar to Lucas et al. [2018] where the value of score s_{it} is the likelihood conditional on the unobservable mixing variable. For that,

$$
\begin{aligned}
s_{it}^{OP} &= \frac{\partial}{\partial f_t} \log p\left(u_t | f_t, \mathcal{F}_t, \theta\right) = \frac{\partial}{\partial f_t} \texttt{log} \int p(u_t, z_t | f_t, \mathcal{F}_t, \theta) dz_t \\
&= \left[\int p(u_t, z_t | f_t, \mathcal{F}_t, \theta) dz\right]^{-1} \int \frac{\partial}{\partial f_t} p(u_t, z_t | f_t, \mathcal{F}_t, \theta) dz_t \\
&= \left[\int p(u_t, z_t | f_t, \mathcal{F}_t, \theta) dz\right]^{-1} \int \frac{\partial}{\partial f_t} p(u_t | z_t, f_t, \mathcal{F}_t, \theta) p(z_t) dz_t \\
&= \left[\int p(u_t, z_t | f_t, \mathcal{F}_t, \theta) dz\right]^{-1} \int \frac{1}{p(u_t | z_t, f_t, \mathcal{F}_t, \theta)} \frac{\partial p(u_t | z_t, f_t, \mathcal{F}_t, \theta)}{\partial f_t} p(u_t | z_t, f_t, \mathcal{F}_t, \theta) p(z_t) dz_t \\
&= \int \frac{\partial \texttt{log} p(u_t | z_t, f_t, \mathcal{F}_t, \theta)}{\partial f_t} \frac{p(u_t | z_t, f_t, \mathcal{F}_t, \theta) p(z_t)}{\int p(u_t, z_t | f_t, \mathcal{F}_t, \theta) dz} dz_t \\
&= \int \frac{\partial \texttt{log} p(u_t | z_t, f_t, \mathcal{F}_t, \theta)}{\partial f_t} p(z_t | u_t, f_t, \mathcal{F}_t, \theta) dz_t \\
&= \mathbf{E}_{z_t}\left[\frac{\partial \texttt{log} p(u_t | z_t, f_t, \mathcal{F}_t, \theta)}{\partial f_t} \middle| \, u_t, f_t, \mathcal{F}_t, \theta\right].
\end{aligned}
$$

Here, the value of standard score s_{it}^{OP} is the expectation of our proposal score s_{it} over z_t, where z_t has the pdf $p(z_t | u_t, f_t, \mathcal{F}_t, \theta)$ distribution.

A.3 Tail dependence for the generalized hyperbolic skew Student-t copula

Consider the bivariate GSt copula. We derive the penultimate tail dependence of a pair of pseudo observables x_{igt} and x_{jgt} in a same group, g, at time t as:

$$\begin{aligned} C_{GSt}(u,u|R_t,\nu_g,\gamma_g) &= F_{MGSt}(F_{GSt}^{-1}(u),F_{GSt}^{-1}(u)|R_t,\nu_g,\gamma_g) \\ &= \int_{-\infty}^{F_{GSt}^{-1}(u)}\int_{-\infty}^{F_{GSt}^{-1}(u)} f_{MGSt}(x_1,x_2|R_t,\nu_g,\gamma_g)dx_1dx_2 \end{aligned} \tag{A.2}$$

where

$$\begin{aligned} f_{MGSt}(x_1,x_2|R_t,\nu_g,\gamma_g) &= c\frac{K_{\frac{\nu+2}{2}}\left(\sqrt{(\nu+Q(x))Q(\gamma)}\right)\exp(x^{'}R_t^{-1}\gamma)}{\sqrt{(\nu+Q(x))Q(\gamma)}^{-\frac{\nu+2}{2}}\left(1+\frac{Q(x)}{\nu}\right)^{\frac{\nu+2}{2}}}, \\ c &= \frac{2^{1-\frac{\nu+2}{2}}}{\Gamma(\frac{\nu}{2})\pi\nu|R_t|^{0.5}}, \\ Q(x) &= x^{'}R_t^{-1}x^{'}, \\ Q(\gamma) &= \gamma^{'}R_t^{-1}\gamma^{'}, \end{aligned} \tag{A.3}$$

and $K_\lambda(.)$ is the modified Bessel function of the third kind with index λ. We obtain $C(u,u|R_t,\nu_g,\gamma_g)$ as the numerical integral. Then, we take the average of $C(u,u|R_t,\nu_g,\gamma_g)$ over T observations for the mean correlation of all assets in the same group. The tail penultimate dependence are,

$$\lambda_L = \sum_{t=1}^{T}\frac{1}{T}\frac{C_{GSt}(u,u|R_t,\nu_g,\gamma_g)}{u}, \text{ and,}$$

$$\lambda_U = \sum_{t=1}^{T}\frac{1}{T}\frac{1-2u+C_{GSt}(1-u,1-u|R_t,\nu_g,\gamma_g)}{u}$$

Table (2.6) reports the penultimate tail dependence of bivariate MGSt copula at $u = 0.005$.

A.4 Posterior inference

From the joint posterior of the dynamic MGSt factor copula model in (2.9), we derive the conditional posterior for each parameters as follows:

$$
\begin{aligned}
p(z_t|u,a,b,f_c,z_{1:(t-1)},\nu,\gamma,\zeta) &\propto \prod_{g=1}^{G}\prod_{i=1}^{n_g}\phi(\tilde{x}_{igt} \mid \rho_{igt}z_t, \sqrt{1-\rho_{igt}^2})\phi(z_t \mid 0,1)\\
p(f_{igc}|u,a,b,z,\nu,\gamma,\zeta) &\propto \prod_{t=1}^{T}\prod_{i=1}^{n_g}\phi(\tilde{x}_{igt} \mid \rho_{igt}z_t, \sqrt{1-\rho_{igt}^2})\\
p(a_g|u,b,f,z,\nu,\gamma,\zeta) &\propto \prod_{t=1}^{T}\prod_{i=1}^{n_g}\phi(\tilde{x}_{igt} \mid \rho_{igt}z_t, \sqrt{1-\rho_{igt}^2})\\
p(b_g|u,a,f,z,\nu,\gamma,\zeta) &\propto \prod_{t=1}^{T}\prod_{i=1}^{n_g}\phi(\tilde{x}_{igt} \mid \rho_{igt}z_t, \sqrt{1-\rho_{igt}^2})\\
p(\nu_g|u,f,a,b,z,\gamma,\zeta) &\propto \prod_{t=1}^{T}\prod_{i=1}^{n_g}\frac{\phi(\tilde{x}_{igt} \mid \rho_{igt}z_t, \sqrt{1-\rho_{igt}^2})}{f_{GSt}(x_{igt}|\nu_g,\gamma_g)}\\
&\quad\times\prod_{t=1}^{T} IG\left(\zeta_{gt}|\frac{\nu_g}{2},\frac{\nu_g}{2}\right) G(\nu_g-4 \mid 2,2.5)\\
p(\gamma_g|u,f,a,b,z,\nu,\zeta) &\propto \prod_{t=1}^{T}\prod_{i=1}^{n_g}\frac{\phi(\tilde{x}_{igt} \mid \rho_{igt}z_t, \sqrt{1-\rho_{igt}^2})}{f_{GSt}(x_{igt}|\nu_g,\gamma_g)}\phi(\gamma_g \mid 0,1)\\
p(\zeta_{gt}|u,f,a,b,z,\nu,\gamma,\zeta_{g,1:(t-1)}) &\propto \prod_{i=1}^{n_g}\frac{\phi(\tilde{x}_{igt} \mid \rho_{igt}z_t, \sqrt{1-\rho_{igt}^2})}{\sqrt{\zeta_{gt}}} IG\left(\zeta_{gt} \mid \frac{\nu_g}{2},\frac{\nu_g}{2}\right)
\end{aligned}
$$

As the conditional posterior of z_t depends on the pseudo observations at time t and the GAS process ρ_{igt}, it is fast to sequentially sample from the Gaussian conjugate distribution. Also, the conditional posteriors of a_g, b_g, ν_g, γ_g and ζ_{gt} only depend on the pseudo observations in group g. Then, we can make parallel inference for $g = 1,\ldots,G$. Finally, conditional on z_t, each time series is independent, for $i = 1,\ldots,n_g$ and $g = 1,\ldots,G$. Then, we also create a parallel estimation procedure for f_{igc}.

A.5 Model selection

The statistics of model selection are calculated based on the average of the log-likelihood. We take the average of the log likelihood after MCMC iterations at the posterior mean of the parameters of interest, $\theta_{int} = \{a, b, f_c, \nu, \gamma\}$, as the integral over the nuisance parameter space, $\theta_{nui} = \{z, \zeta\}$. Then, the AIC, BIC and DIC are given,

$$
\begin{aligned}
AIC &= -2\mathbb{E}_{\theta_{nui}}\left[\log\left(p\left(u|\bar{\theta}_{int}, f, \mathcal{F}\right)\right)\right] + 2k \\
BIC &= -2\mathbb{E}_{\theta_{nui}}\left[\log\left(p\left(u|\bar{\theta}_{int}, f, \mathcal{F}\right)\right)\right] + k\log T \\
DIC &= -4\mathbb{E}_{\theta}\left[\log p\left(u|\theta, f, \mathcal{F}\right)|u\right] + 2E_{\theta_{nui}}\left[\log\left(p\left(u|\bar{\theta}_{int}, f, \mathcal{F}\right)\right)\right]
\end{aligned}
$$

where k is the number of parameters of interest.

Appendix B

Appendix of Chapter 3

B.1 The step size

At iteration i, the gradient vector $g^{(i)} = \nabla_{\lambda^{(i)}} \texttt{ELBO}$ and step size vector $\varrho^{(i)}$ are used to update the parameters of copula model. The idea is to downscale the gradient by the exponential moving average of the second moment of the gradient. The parameter with a large derivative will have a quickly decline learning rate. We have for each element $\varrho_j^{(i)}$, where $j = 1, \ldots, N$,

$$\begin{aligned}
\varrho_j^{(i)} &= i^{-0.5+\epsilon} \times \left(1 + \sqrt{s_j^{(i)}}\right)^{-1} \\
s_j^{(i)} &= 0.1 g_j^{2(i)} + 0.9 s_j^{(i-1)}
\end{aligned}$$

where $\epsilon = 10^{-16}$ to guarantee the step size decay, and an initialization $s_j^{(1)} = g_j^{2(1)}$ at $i = 1$.

Appendix C

Appendix of Chapter 4

C.1 Empirical illustration

Figure C.1 in the Appendix shows the posterior means of model parameters in the truncated factor vine copula. The histogram of the common latent variable is plotted in the top left corner. The figure on the top right and bottom right show the histogram the Kendall-τ correlation in the truncated factor vine copula at the first tree and second tree. The histogram of degree of freedom ν is on top right corner. Note that, $\nu = 0$ represents non Student-t bivariate copulas.

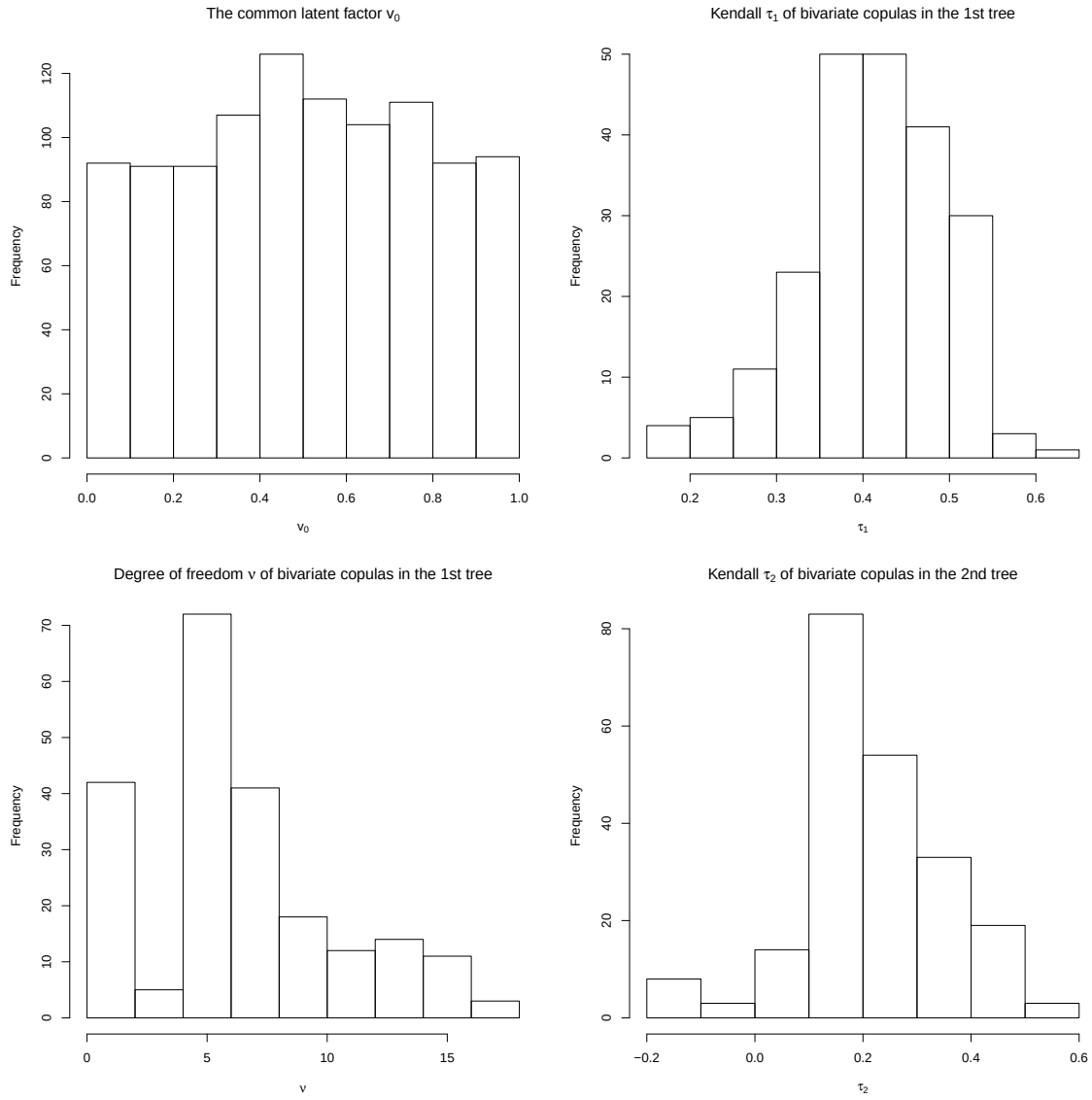

Figure C.1: Histogram of the Kendall-τ correlation and degree of freedom ν of bivariate copulas in stock return data.

The figure on the top left shows the histogram of the common latent variable estimated at the posterior means. The figure on the top right and bottom right show the histogram the Kendall-τ correlation in the truncated factor vine copula at the first tree and second tree. The histogram of degree of freedom ν is on top right corner. Note that, $\nu = 0$ represents non Student-t bivariate copulas.